Characterization of the facial phenotype associated with fetal alcohol syndrome using stereo-photogrammetry and geometric morphometrics

Tinashe E.M. Mutsvangwa

Biomedical Engineering

University of Cape Town

A thesis submitted for the degree of

Doctor of Philosophy

May 2009

I would like to dedicate this thesis to my family for all the support and love they have given me throughout my academic career. I would also like to dedicate this to Dominique Reinecke for all her words of encouragement during moments of acute frustration.

Acknowledgements

I would like to acknowledge first and foremost my supervisor Dr Tania Douglas. Her constant advice and guidance has been critical in determining my career path from my masters degree to the present body of work. I would also like to thank to my co-supervisor Ernesta Meintjes for all her assistance in this project.

Special acknowledgement goes to the following people for their help in the design and construction stage of the project:

- Dr Julian Smit for his tutorials on stereo-photogrammetry.
- Charles Harris for constructing the stereo-imaging tool from the design plans.
- Iekraam Fakier for his assistance with the trigger mechanism for the cameras.

Thanks go to the following people from the Foundation of Alcohol Related Research (FARR) for the their assistance during the data gathering stage of this project:

- Professor Dennis Viljoen
- Leana Olivier
- Candice Chetty
- Leigh-Anne Fourie

Gratitude goes out to all the friends, especially people I have met during my time at UCT. Last but not least I would like to thank my family for their support and patience throughout my entire academic career.

Abstract

Fetal Alcohol Syndrome (FAS) is a clinical condition caused by excessive prenatal alcohol exposure and is regarded as a leading identifiable and preventable cause of mental retardation in the Western world. The highest prevalence of FAS was reported in the wine-growing regions of South Africa but data for the rest of the country is not available. Required, therefore, are large-scale screening and surveillance programmes to be conducted in South Africa in order for the epidemiology of the disease to be understood. Efforts to this end have been stymied by the cost and labour-intensive nature of collecting the facial anthropometric data useful in FAS diagnosis.

Stereo-photogrammetry provides a low cost, easy to use and non-invasive alternative to traditional facial anthropometry. The design and implementation of a landmark-based stereo-photogrammetry system to obtain 3D facial information for fetal alcohol syndrome diagnosis (FAS) is described. The system consists of three high resolution digital cameras resting on a purpose-built stand and a control frame which surrounds the subject's head during imaging.

Reliability and assessments of accuracy for the stereo-photogrammetric tool are presented using 275 inter-landmark distance comparisons between the system and direct anthropometry using a doll. These showed the system to be highly reliable and precise. Stereo-photogrammetry showed sub millimetre error in precision in landmark placement for all landmarks on the doll. In addition 100% of the inter-landmark distances were within a 1.5 mm error range and 92.36% within a 1 mm error range when direct anthropometry and stereo-photogrammetry were compared. An investigation into the effects of pose using two sets of images of 5 subjects showed the stereo-photogrammetric system to be highly reliable, with an average 72.25% of the measured distances within a 1 mm error range.

Comparisons, using human subjects, between the imaging tool and a dysmorphologist and between the imaging tool and a commercially available system

are also reported; these showed mixed results. Data for these two analyses, however, were taken during screening exercises for FAS, and subject pose was not controlled for during measurements.

Landmark-based morphometric analysis holds promise for quantitative assessment of 3D craniofacial data. An application of facial shape analysis to characterize the facial anomalies associated with fetal alcohol syndrome (FAS) in a mixed ancestry population using two approaches is presented. One uses generalized Procrustes analysis and principal component analysis, applied to stereo-photogrammetrically derived 3D coordinates of landmarks taken from thirty-four subjects (n=17 FAS and n=17 normal controls). The other approach uses generalized Procrustes analysis, regression and discriminant function analysis on the same data.

Comparisons of the FAS and control facial shapes at age 5, at age 12 and for both groups combined, revealed that the FAS face is characterized by small palpebral fissures, thin upper lips and midfacial hypoplasia. Classification of subjects using leave-one-out cross-validation showed that the five-year old group could be classified with higher accuracy than the twelve-year old group. Finally, comparisons of the FAS faces and of the normal faces at different ages were made using only the second morphometric approach. Greater differences in facial shape were found between FAS individuals at different ages than for control individuals, supporting the notion that FAS facial anomalies disappear with age.

The stereo-photogrammetry system and facial shape analysis methods presented show potential for large scale screening and surveillance studies for FAS.

Contents

List of Figures

Nomenclature

Roman Symbols

$PC(r)$ Principal component r

S_i^{PR} Procrustes residuals coordinates

x_0, y_0 Position of the principal point relative to the reference system fixed within the image plane

X_0, Y_0, Z_0 Exterior orientation position parameters

S_i^P Procrustes fit coordinates

C Perspective centre

L_1, ..., L_{11} DLT parameters

r_{11}, ..., L_{33} Rotation matrix elements

x, y 2D image space coordinate system

X, Y, Z 3D object space coordinate system

Greek Symbols

Δx, Δy lens distortion parameters

ω, φ, κ Exterior orientation rotation parameters

$\overline{\alpha}$ Procrustes mean shape

Chapter 1

Introduction

Fetal Alcohol Syndrome (FAS) is a clinical condition caused by excessive maternal consumption of alcohol during pregnancy (Burd *et al.*, 2003). It is widely regarded as a leading identifiable and preventable cause of mental retardation and neurological deficit in the Western world. The highest prevalence of FAS worldwide was reported among first-grade children in a wine-growing region in the Western Cape province of South Africa at approximately 68.0 - 89.2 per 1000 children (May *et al.*, 2007). The frequency of occurrence of the condition in other regions of South Africa is not known and must be determined. It is imperative that large-scale screening and surveillance programmes be conducted in order for the epidemiology of the syndrome to be understood and effective intervention measures to be introduced. This will require cheap, fast and non-labour intensive methods, because South Africa has other competing health issues including HIV/AIDS and tuberculosis.

Diagnosis of FAS depends on evidence of growth retardation, CNS neurodevelopmental abnormalities and a characteristic pattern of facial anomalies, specifically short palpebral fissure length, smooth philtrum, flat upper lip and flat midface, in addition to a positive history of the patient's mother having consumed alcohol during pregnancy (Hoyme *et al.*, 2005). The facial phenotype associated with FAS has been emphasized in clinical diagnosis (Clarren *et al.*, 1987). Traditionally, diagnosis of the facial anomalies associated with FAS used the "gestalt" method. In the hands of trained dysmorphologists, this method can be sufficiently accurate and reproducible for diagnosing the condition (Clarren *et al.*, 1987). If misdiagnosis occurs, however, this adversely affects the patient through stigmatisation and can severely stifle any screening and prevalence efforts that could help track the disorder (Astley & Clarren, 1995). A more accurate and objective method for FAS diagnosis

will therefore be useful. Attempts have been made to standardize the diagnosis of the FAS facial features by introducing a quantitative case definition derived both from direct anthropometric and indirect photogrammetric methods (Astley & Clarren, 1995, 1996, 2000; Moore *et al.*, 2001, 2002).

Current methods of identifying the FAS facial phenotype rely on linear distances, neglecting the shape information contained in the relative positions of the facial landmarks that define these distances. Statistical shape analysis has been used in the identification of syndromes that display facial dysmorphology and holds promise in the characterisation of the FAS facial phenotype (Mutsvangwa & Douglas, 2007).

This thesis explores the use of a cheap, non-labour intensive, but accurate imaging system that produces three-dimensional (3D) facial information, together with objective analytical methods to aid in diagnosis of and screening for FAS. A detailed assessment of reliability and precision of the imaging tool is presented by means of comparisons of repeated measurements on an annotated doll, and comparisons of the imaging tool with a dysmorphologist and with the commercially available Vectra 3D system in obtaining inter-landmark distance measurements on human subjects. Two methods of facial shape analysis using 3D landmark data obtained using the imaging tool are presented. The two methods are compared in their characterization of the FAS facial shape. Both methods are used to provide some insight into the FAS facial phenotype at different ages by comparing the FAS and the normal facial shapes at five and at twelve years of age. The notion that the FAS facial anomalies disappear with age, is tested using shape analysis. Specific methodologies for the consideration of extrinsic factors of shape variability such as age and size are applied. Classification of subjects using pattern recognition techniques is demonstrated.

Approval for the component of the work involving human subjects was obtained from the Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town. Informed written consent was obtained from the parents/guardians of all subjects.

1.1 Objectives

The objectives of the project described here were to demonstrate the use of an inexpensive stereo-photogrammetric imaging tool to obtain the 3D coordinates of facial landmarks in children and to use these landmarks to characterize the FAS facial phenotype in a sample population by way of geometric morphometric methods. The following steps were taken to achieve these objectives:

1. Designing and constructing a cheap, portable, easy-to-use stereo-photogrammetric imaging tool capable of imaging large numbers of subjects at different ages, including infants.

2. Testing of the system for reliability and precision using a doll and human subjects.

3. Capturing images of subjects with FAS and controls in the Northern Cape regions of South Africa.

4. Using facial shape analysis to characterize the FAS facial phenotype for the study population.

5. Classification of subjects into two groups, FAS and normal, on the basis of their facial shape, using pattern recognition algorithms.

1.2 Thesis outline

Chapter 2 begins with a background to FAS, the current method of diagnosis and prevalence statistics of the syndrome. Alternative anthropometric methods used in obtaining data for assessing the FAS facial anomalies are then described including photography, stereo-photogrammetry and laser scanning. Statistical techniques used in analysing anthropometric data are presented and they include, linear-based methods, Likert scales, pattern profiling and geometric morphometrics. The chapter concludes with a brief description of the various diagnostic criteria used in clinical assessment of FAS.

Chapter 3 begins with a description of the theoretical framework underlying stereo-photogrammetry. The two main methods used are the direct linear transform and bundle adjustment. Practical considerations in designing a stereo-photogrammetry imaging tool are listed before the specific imaging task required from the imaging tool is presented. Next, the process of developing the imaging tool is described, including calibration of the frame and cameras and a description of system components. Preliminary system tests are reported.

Chapter 4 begins with an overview of how accuracy and reliability have been assessed in both direct and indirect anthropometry. Studies comparing anthropometric measurements obtained from a doll and from children using direct anthropometry and stereo-photogrammetry are then reported. Two indirect measurement systems, namely static

stereo-photogrammetry and the commercial Vectra 3D system are then compared. A discussion of the general performance of the stereo-photogrammetry method concludes the chapter.

Chapter 5 is concerned with the application of geometric morphometrics in characterising the FAS facial shape. The chapter begins with some theoretical background on facial shape analysis. Two approaches to facial shape analysis with respect to FAS are presented. One uses principal component analysis with the application of pattern recognition algorithms for classification and the other uses regression for size and age correction, before the use of discriminant function analysis for classification. The results of both approaches are reported, compared and discussed.

Chapter 6 provides a final discussion of the work presented. Conclusions are drawn and recommendations for future work are made.

1.3 Publication

Some of the work covered in Chapters 3 and 4 has been accepted for publication as follows:

Mutsvangwa TEM, Smit J, Hoyme EH, Kalberg W, Viljoen DL, Meintjes EM, Douglas TS. Design, construction and testing of a stereo-photogrammetric tool for the diagnosis of fetal alcohol syndrome in infants. *IEEE Transactions on Medical Imaging*, in press.

The author of this thesis was responsible for the design and testing of the system described in the above publication, as well as for data collection, statistical analysis and manuscript preparation.

Chapter 2

Background

2.1 Fetal Alcohol Syndrome

A syndrome is a recognizable pattern of malformations, and/or disruptions occurring together to characterize a particular disease (Hunter, 2002). FAS is the most severe syndrome of the range that falls under the umbrella term fetal alcohol spectrum disorders (FASD) (Astley, 2006) and is widely regarded as a leading identifiable and preventable cause of mental retardation and neurological deficit in the Western world (Burd *et al.*, 2003). It is recognized as a major public health concern in most societies worldwide and is known to affect all racial and ethnic groups (Manning & Hoyme, 2007). Jones & Smith (1973) set forth diagnostic criteria for FAS and this definition has changed very little over the past three decades. Specifically, FAS is defined by prenatal and/or postnatal growth retardation, central nervous system dysfunction, reflected in mental and behavioural deficits and a unique collection of minor facial anomalies (Astley & Clarren, 1995; Burd *et al.*, 2003; Clarren *et al.*, 1987; Jones, 1986; Jones & Smith, 1973; Manning & Hoyme, 2007; May *et al.*, 2000, 2007). Prenatal alcohol exposure does not always result in FAS. Previously, the term Fetal Alcohol Effects (FAE), first used in 1978 (Sampson *et al.*, 1997), was used to refer to behavioural and cognitive problems in individuals with prenatal alcohol exposure but without full-blown FAS (Astley & Clarren, 2000). The term has been abandoned in the mainstream due to its broad definition which brackets, for example, individuals who were exposed prenatally to alcohol and exhibit mild cognitive disabilities with individuals who may have been exposed to heavy doses of alcohol in the womb and present severe signs of FAS but no facial anomalies associated with the syndrome. In addition, Astley & Clarren (2000) argue that the use of the term, "alcohol effect",

inappropriately implies a certifiable causality between alcohol exposure and the outcome. The Institute of Medicine (IOM 1996), described three other outcomes from fetal alcohol exposure. Partial fetal alcohol syndrome (PFAS) includes the FAS facial anomalies, but only one of the following: growth retardation, structural brain deformities and cognitive abnormalities. Alcohol related birth defects (ARBD) require confirmed maternal exposure to alcohol, the FAS facial phenotype and more than two congenital structural defects relating to other organ systems in the body. The fourth outcome is alcohol-related neuro-developmental disorders (ARND), which requires confirmed maternal alcohol exposure and at least one of the following: structural brain abnormalities, microcephaly or evidence of behavioural or cognitive inconsistencies (Manning & Hoyme, 2007). Partial fetal alcohol syndrome, ARBD and ARND can be considered "less severe" manifestations of FAS and could also relate to the timing of alcohol exposure during pregnancy, the genetic factors in the mother or fetus affecting alcohol metabolism or susceptibility, and the dose of alcohol consumed, all of which can modify the outcome of exposure (Sampson *et al.*, 1997). Differentiating between FAS and the partial forms of the syndrome is difficult since criteria overlap and the physical and behavioural expression can be widely variable among individuals (Burd *et al.*, 2003).

The effects of FAS are not reversible. Individuals with the syndrome have to live with life-long disabilities. Compounding the primary disabilities are secondary effects including diminished self esteem, depression, a reduced capacity to integrate socially, school failure because of the associated neurological deficits, and social chastisement from society (Astley & Clarren, 1995). They may display inappropriate sexual behaviour, and often end up in juvenile detention should the syndrome go unnoticed (Chudley *et al.*, 2005). The secondary effects of FAS are costly for the individual, family and society at large. The potential of cost saving from early detection and intervention is high (Astley & Clarren, 1995). In addition, prevention has the capacity to save money (in the United States, the cost of care for FAS is US $1.4-2.0 million per lifetime per case) (Klug & Burd, 2003). Early detection and prevention of secondary disabilities depends on accurate and reliable diagnostic aids and screening tools (Astley & Clarren, 1996). The failure to diagnose FAS is a consequence of the difficulty of diagnosis. Growth retardation and neurological disorders can be accurately and reliably determined although neither nor both combined are specific to FAS. Although brain damage is the most significant disability for individuals exposed to prenatal alcohol exposure (Astley & Clarren, 2000), the unique pattern of facial anomalies is the only feature from the diagnostic definition of FAS that is specific to the condition (Astley & Clarren, 1995; Clarren *et al.*, 1987; Jones & Smith,

1973). The case definitions for FAS diagnosis have therefore focused on the unique cluster of features.

2.1.1 Prevalence

The prevalence of FAS in the developed world has been estimated at 0.97 per 1000 live births (May *et al.*, 2007). In the United States the estimate is between 0.33 and 2 per 1000 (Burd *et al.*, 2003). The highest prevalence reports for FAS worldwide have been from the Western Cape Region of South Africa, with 40.5-46.4 cases per 1000 first grade children reported by May *et al.* (2000), 65.2 - 74.2 per 1000 by Viljoen *et al.* (2005) and 68.0 - 89.2 per 1000 by May *et al.* (2007). These figures pertain to full-blown FAS, which is only one of the outcomes of prenatal alcohol exposure. Prevalence of FASD is expected to be higher with some estimates at 9-10 per 100 (Manning & Hoyme, 2007). Several studies have been done on the prevalence in South Africa (Croxford & Viljoen, 1999; May *et al.*, 2000, 2004, 2005, 2007, 2008; Viljoen *et al.*, 2002, 2005). The focus has been in urban areas and the wine-growing regions of the Cape province. Other regions of South Africa have not been investigated. Understanding the epidemiology of the syndrome in this country requires large-scale screening and surveillance programmes, which would allow effective intervention measures to be introduced. However, because of South Africa's other competing health issues, namely HIV/AIDS and tuberculosis, resources and expertise available for such programmes are limited.

2.1.2 Craniofacial measurements

Traditionally, diagnosis of FAS from facial features used the "gestalt" method. In the context of FAS diagnosis, this refers to a method where the whole face is considered as a unit and not just the sum of its parts. The gestalt method, an intrinsically qualitative approach, is considered an acceptable standard of syndrome diagnosis that can, when used by a trained professional, be sufficiently accurate and reproducible (Astley & Clarren, 1995). When the gestalt method is used for a syndrome such as FAS, however, which can present in subtle forms and is influenced by ethnic variations (Moore *et al.*, 2007), accuracy may come into question, especially when such an approach is used by less trained individuals. In some areas of the world, diagnosis is sometimes performed by general practitioners or paediatricians as opposed to trained dysmorphologists. This can lead to diagnostic inaccuracy which in turn curtails the screening and surveillance efforts of the syndrome (Astley & Clarren, 1995). A more objective, quantitative approach to FAS

diagnosis, using facial features, has been the goal of several research groups over the past decade (Astley, 2004; Astley & Clarren, 1995, 1996, 2000; Clarren *et al.*, 1987; Fang *et al.*, 2008; Huang *et al.*, 2005; Meintjes *et al.*, 2002; Moore *et al.*, 2001, 2002, 2007; Mutsvangwa & Douglas, 2007). Sampson *et al.* (1997) lists the facial anomalies that are typical in FAS. They can be divided into two categories: descriptive anomalies which include short palpebral fissures, midfacial hypoplasia, smooth philtrum and a thin upper lip; and associated features which include epicanthal folds, low nasal bridge, minor ear anomalies, short noses and micronathia.

2.2 Direct anthropometry

Quantitatively delineating the FAS facial phenotype requires measurements of the facial anomalies associated with the condition. Anthropometry is defined as the biological science of measuring size, weight and proportions of the human body (Farkas, 1994), especially on a comparative basis. Direct anthropometry relates to measurements taken from the human body using physical measurement devices and usually involves some level of contact with the parts of the body being measured. Lengths, angles and circumferences of body features lend themselves to direct anthropometric measurement. Instruments used in direct anthropometry include hand-held callipers, rulers, cloth tapes and protractors. In the FAS context, direct anthropometry has been and is still used to measure eye distances (inner canthal, outer canthal, interpupillary and palpebral fissure lengths), philtrum length, head circumference, etc., which are compared with population norms (Astley & Clarren, 1995; Moore *et al.*, 2002). These measurements are centred around the craniofacial area and generally require intimate contact between the face and the measuring instrument. In vulnerable areas such as the eyes this can be potentially hazardous, especially when taking measurements of infants or syndromic patients. In addition to the risk of injury, there is also the risk of measurement error due to parallax and, because measurements can be time-consuming, investigator fatigue and patient discomfort. Some facial anomalies considered part of the FAS facial phenotype, such as philtrum smoothness, do not lend themselves to measurement using physical instruments. Advantages of taking measurements from the body using direct anthropometry are the relative ease of use and low cost of the measuring instruments. Direct anthropometry also offers the ability to take measurements of areas of the head and face covered by hair and allows measurements that require special techniques (e.g. taking measurements of bony landmarks after compression of the overlying soft tissue) (Moore *et al.*, 2002).

2.2.1 Using direct anthropometry to standardise FAS diagnosis

Astley & Clarren (1995) used a combination of direct anthropometry and 3-point Lickert scale ratings (see Figure 2.1) of some features in their proposed FAS screening tool. The purpose of their study was to derive, by comparison with the then gestalt gold standard, a multivariate, quantitative case definition for the FAS facial phenotype. Their study population consisted of mixed race groups (0.2-10 years of age), 194 subjects (39 FAS and 155 controls) in total. They took several morphometric measurements including those from the eyes, midface and mouth, together with weight, height and occipital frontal circumference to make a total of 13 physical and facial measures. The study population was split into two and one group (20 FAS and 77 controls) was used to create a list of the most differentiating features between FAS subjects and normal controls using discriminant function analysis. The discriminant function found using data from the first group was applied to the second group (19 FAS and 78 controls). The discriminant function identified thin upper lip, smooth philtrum and small palpebral fissures as the three features that best discriminated between children with and without FAS. A classification of the first group using the discriminant function resulted in 100% sensitivity and 90% specificity. Sensitivity is defined as the proportion of true positives correctly identified as positive and specificity is defined as the proportion of true negatives correctly identified as negative. Accuracy is the proportion of true results (both positive and negative) from the sample. Application of the function to classification of the second group resulted in 100% sensitivity and 87.3 % specificity. The study sucessfully identified which features were more relevant in differentiating FAS and normal subjects. The study did not, however, reveal the extent of deviation from the norm for which a feature would become indicative of FAS.

Moore *et al.* (2001) used direct anthropometry to take 21 craniofacial measurements of individuals exposed prenatally to alcohol and control subjects, in order to obtain an objective, multivariate case definition of FAS and partial fetal alcohol syndrome (PFAS). They used discriminant analysis with stepwise variable selection to find the subset of craniofacial measurements that best differentiated the three groups (FAS, PFAS and controls). Discriminant analysis identified two craniofacial measurements (head circumference and bigonial breadth) that could distinguish between FAS and controls with 100% accuracy. They also found that five measurments (head circumference, bigonial breadth, midfacial depth, byzygomatic breadth and maxillary arc) could differentiate children with FAS from

children with PFAS with 88% accuracy. More importantly, discriminant analysis identified six craniofacial measurements (minimal frontal breadth, bigonial breadth, midfacial depth, palpebral fissure length, head circumference and maxillary arc) that could distinguish between alcohol exposed and non-alcohol exposed subjects with 98% sensitivity and 90% specificity. Their study demonstrated the use of objective measurements in extending the use of the facial phenotype from identifying individuals with the most obvious facial disruptions to including individuals with moderate expressions of the facial anomalies due to prenatal alcohol exposure.

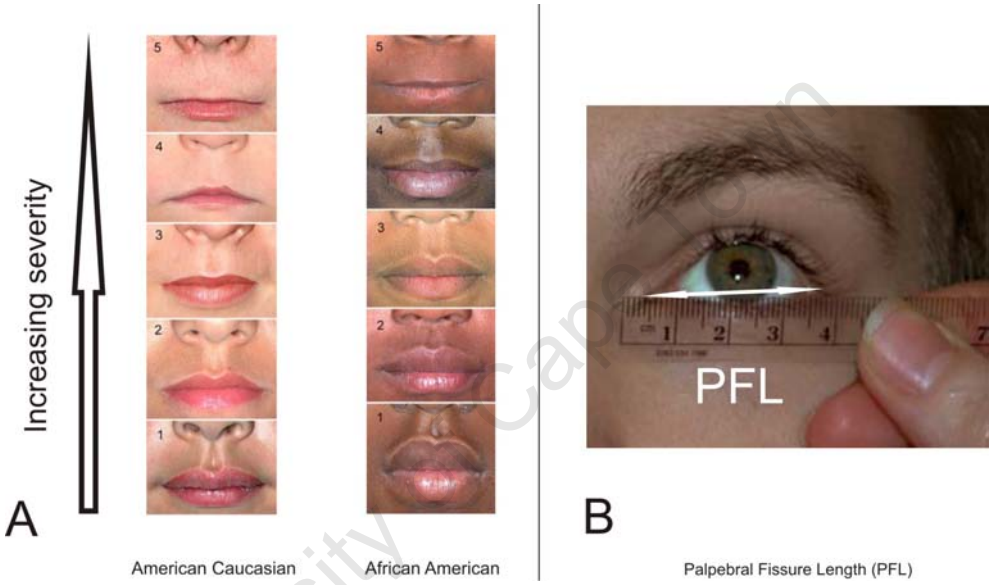

Figure 2.1: A: Philtrum Lickert scales for two ethnic groups, Caucasian and African American. and B: A palpebral fissure length measurement using a clear hand-held ruler. (Adopted and modified with permission from Astley (2004)).

In a later study, Moore *et al.* (2002) used 21 anthropometric measurements from 100 prenatally alcohol exposed and 31 control subjects standardized to age- and sex-matched population norms. They then employed multivariate statistical methods on the normalized data to assess for: significant craniofacial measurements differentiating three groups (FAS, PFAS and controls), and the effects of race and age on diagnostic classifications; they finally created a pattern profile for each group which they used to determine the degree of similarity between groups. Their results showed the pattern profile of the FAS sample to be significantly smaller than the PFAS group and the reference controls for most craniofacial measurements. In addition, the PFAS group was intermediate in size

between the reference and the FAS group. They also reported that there existed a unique pattern profile seen in the alcohol exposed groups (both FAS and PFAS) which was not only related to size, but facial shape.

2.3 Indirect anthropometry

2.3.1 Photography

Measurements from photographs offer several advantages over direct anthropometry. The outlines of a photograph do not move during a measurement, in contrast with direct measurements. Obtaining photographs is generally faster than taking measurements directly. The measurements can be repeated, if necessary, from photographs, whereas repeated direct measurements of the subject may not be possible. Photographs, however, also have disadvantages compared to direct anthropometry. Photographs need some measure of scale, otherwise only ratios of measurements can be compared. They are also vulnerable to lighting conditions and photographic distortion during imaging. The 2D nature of photographs makes it impossible to measure tangential arcs. For the best results the subject is required to be photographed with landmarks marked on the skin and a standardized positioning of all images, both of which may not always be possible. Guidelines exist however, for both image acquisition and taking measurements (Farkas, 1994).

2.3.1.1 Photography in FAS diagnosis

Clarren *et al.* (1987) used photographs in assessing the facial manifestations of fetal alcohol exposure. Copies of full face and lateral facial photographs of 42 subjects (twenty one 7-year old children exposed to heavy quantities of alcohol and twenty one 7-year old children with negligible alcohol exposure) were sent to each of seven clinical experts to be screened for a FAS-related facial appearance. In addition, newly introduced morphometric methods were used to analyse the same photographs for facial shape characteristics that differentiated the two groups of subjects from each other. The morphometric method they employed used 2D landmarks digitized from the images. The analysis used triangles defined by sets of three landmarks. Mean shapes of these triangles were compared between the alcohol exposed and the non-alcohol exposed subjects and the results confirmed the facial phenotype identified by the experts; the facial phenotype consisted of short palpebral fissures, a relatively long and flat midface and a retrusive mandible. This

was the first reported use of shape analysis methods to describe the facial appearance associated with fetal alcohol exposure.

In a follow-up to their previous work on defining a screening tool for FAS diagnosis (Astley & Clarren, 1995), Astley & Clarren (1996) employed anthropometric measurements from photographs, consisting of: one ocular measurement, namely a reduced palpebral fissure length/inner canthal distance ratio, the phenotypic expression of philtrum smoothness and upper lip thinness as recorded from a five-point Likert scale, upper lip thinness as measured by an objective metric of shape called circularity, and philtrum smoothness as measured by an objective metric called pixel luminosity. Their study population consisted of 42 subjects (0 to 27 years of age) with FAS and 84 controls, age-matched without FAS. They split the data set into two equal groups of 63 members each matched on age, gender and race and used the first group to identify the best features that differentiated individuals with and without FAS. The second group was used for cross-validation. Using stepwise discriminant analysis they identified three facial features (reduced palpebral fissure length/inner canthal distance ratio, smooth philtrum measured on the Likert scale, and thin upper lip measured on a continuous scale) as the cluster of features that differentiated individuals with and without FAS with 100% accuracy for both groups. This study illustrated that FAS diagnosis could be performed accurately using objective measurements of facial anomalies only and ultimately led to the establishment of the widely adopted 4-Digit Diagnostic Code (Astley & Clarren, 2000).

The 4-Digit Diagnostic code uses a coding system and images to quantify the magnitude of expression of four key diagnostic features in the following order (1) growth defiency; (2) the FAS facial phenotype; (3) brain dysfunction; (4) gestational alcohol exposure (Astley, 2006; Astley & Clarren, 2000). The magnitude of expression of each feature is ranked independently on a four-point Likert scale with 1 reflecting the complete absence of the FAS feature and 4 reflecting a strong "classic" presence of the FAS feature. Nine unique diagnostic categories ranging from "no cognitive/behavioural or sentinel findings detected" to "FAS" can be used to diagnose individuals of all ages and races who present across the full spectrum of alcohol exposures and outcomes. This method was developed using medical/research records of 1014 patients diagnosed at the Washington State Fetal Alcohol Syndrome and Diagnostic and Prevention Network in the United States. Typically, a FAS diagnosis requires ranks of 3 or 4 in all the four diagnostic feature categories. The use of the diagnostic code was compared to a traditional diagnosis for the first 454 patients from the 1014; the diagnostic categories were FAS, atypical FAS

(AFAS) or possible fetal alcohol effects (PFAE). The outcomes were more accurately and comprehensively documented by the new method (Astley & Clarren, 2000).

Huang *et al.* (2005) note that photographic measurements are limited to point-to-point or curvature-based geometrical features, while facial pattern analysis has no such restrictions. Huang *et al.* (2005) developed an automated feature selection algorithm on a training set that applied principal component analysis to project the face image from a 2D frontal photograph to a lower dimensional face subspace. Multidimensional linear discriminant analysis was then used for classification of new subjects. In general, pattern recognition methods attempt to find the best subset of features that can discriminate between two groups. The approach employed by Huang *et al.* (2005) achieved modest results (a FAS identification sensitivity of 59% and specificity of 62%). It did, however, show the potential to categorize faces with respect to the presence of FAS without making assumptions about the selection of facial features to be measured.

2.3.2 Stereo-photogrammetry

Static stereo-photogrammetry refers to the special case where 3D information of a scene can be extracted from static 2D images (Douglas, 2004). This form of photogrammetry allows a rapid and inexpensive way of acquiring 3D information. It offers all the advantages of photography over direct anthropometry with the added advantage of acquisition of 3D data. This in turn allows more sophisticated data analysis, better able to deal with the inherent 3D nature of the human form. The main disadvantage of static photogrammetry is that only 3D landmarks lying on boundaries of a structure (for example the corners of the mouth) can easily be identified. Points on smooth surfaces, e.g. on the forehead, cheeks and chin, are not easily measurable (Farkas, 1994).

2.3.2.1 Stereo-photogrammetry in FAS diagnosis: distance measurements

Static stereo-photogrammetry has been used as an alternative to direct anthropometry in diagnosing the FAS facial phenotype (Douglas *et al.*, 2003b; Grobbelaar & Douglas, 2007; Meintjes *et al.*, 2002; Mutsvangwa & Douglas, 2007). Meintjes *et al.* (2002) developed a stereo-photogrammetry imaging tool to measure facial dysmorphology in children. The tool consisted of a control frame comprising a head-and a chin-rest with vertical supports on each side (the tool is shown in Figure 2.2). The tool was designed with 11 well distributed retro-reflective control markers. The 3D coordinates of these makers were known, and were used in the calibration of the images. Among the typical facial anomalies of FAS,

several occur in the eye region; inner-canthal, outer canthal and interpupillary distances of subjects were measured from images taken using the imaging tool. These measurements were then compared with measurements that were performed by dysmorphologists in the conventional manner, i.e. using direct anthropometry. No statistically significant difference between palpebral fissure lengths obtained using the direct and indirect modalities were found. Discrepancies in the inner canthal and interpupillary distances were attributed to parallax and possible eye movement of subjects during clinical evaluation.

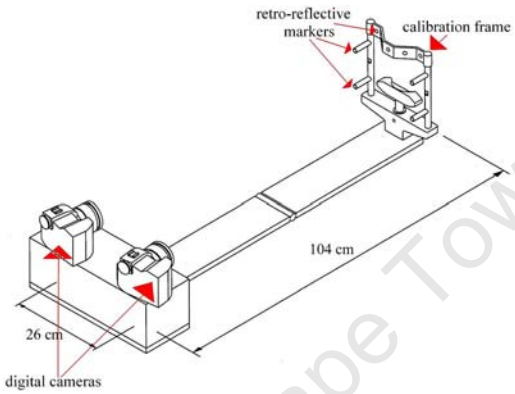

Figure 2.2: A static stereo-photogrammetry imaging tool.

In a follow-up study, automated eye feature extraction was attempted using images of 7-year old subjects taken using the imaging tool (Douglas *et al.*, 2003b). Measurements obtained using an automatic eye extraction algorithm were compared to those taken manually from the images for four eye distances (palpebral fissure length, inner-canthal, outer-canthal and interpupillary distances). Average absolute differences between the automated and manual methods were less than 1mm for palpebral fissure length and interpupillary distance. The difference between palpebral fissure lengths measured using the two methods was less than 1mm for 80.4% of the subjects, while 100% of the subjects had a less than 1mm difference for interpupillary distance. Inner-canthal and outer-canthal distances did not compare favourably. The reason the authors proferred for the differences in measurements between the two methods, was that landmarks defining the inner- and outer canthal distances were selected at different depths in the stereo images. They also suggested that another likely source of error was the innaccuracy in automated selection of corrresponding points between the stereo images.

The role of depth in eye distance measurements (Douglas *et al.*, 2003c) and in eye distance and lip measurements (Douglas *et al.*, 2003a) was assessed by comparing mea-

surements obtained using single- and stereo-photogrammetry. Using images obtained from the above stereo-photogrammetry imaging tool, an operator took measurements of the eyes (inner-canthal, outer canthal and interpupillary distances) and lips (upper lip width and upper lip height). These measurements are considered relevant in FAS diagnosis. The images were taken from single and stereo photographs. In both cases they reported no differences between measurements in 2D (ignoring depth) and 3D in an idealized system where the real-world coordinates of points were known from 3D calibration of stereo photographs. However, differences were found between measurements taken from single frontal photographs and those obtained using stereo-photogrammetry. They concluded that measurements from single frontal photographs are prone to errors due to misalignment of the camera, the face and the calibration frame.

2.3.2.2 Stereo-photogrammetry in FAS diagnosis: facial shape analysis

Multivariate statistical procedures applied to distances, angles and distance ratios have been the standard way of assessment of facial anomalies associated with FAS (Astley & Clarren, 1995, 1996; Moore *et al.*, 2001, 2002, 2007; Naidoo *et al.*, 2006) regardless of data acquisition modality. The limitations of using these measurements in anthropometric analyses have been well documented (Slice, 2005). Primarily, the limited set of linear distances, ratios or angles often fail to capture the entire spatial arrangement between the landmarks from which they are derived. Analysis of the FAS facial phenotype by using geometric morphometrics applied to 3D landmarks obtained with static stereo-photogrammetry was demonstrated by Mutsvangwa & Douglas (2007). Geometric morphometric methods retain all geometric information while simultaneously providing a statistical platform to investigate biological organisms in a quantitative way (Slice, 2005). In addition, the results of the analyses can be readily related back to the physical space of the specimens and graphically viewed.

Mutsvangwa & Douglas (2007) used Procrustes analysis, principal component analysis and discriminant analysis (geometric morphometric methods are described in detail in Chapter 5) on 3D landmarks obtained from images using the stereo-photogrammetry imaging tool described above. Study data were obtained during the screening of first-grade children from disadvantaged communities in the Gauteng and Northern Cape Provinces of South Africa for FAS. Twenty normal and 14 FAS subjects were used in the analysis. The landmarks included those lying on features reported to be the most significant indicators of the FAS facial phenotype and used in diagnosis, namely palpebral fissures, the

philtrum and the upper lip and landmarks in the mid-facial region that may be affected by midfacial hypoplasia in FAS subjects (see Figure 2.3). The 3D landmarks were used to compare the facial shapes defined by features associated with FAS in subjects with FAS and normal controls in an attempt to characterize the facial phenotype associated with FAS. The analysis revealed significant differences in facial shape between the two groups, broadly confirming the FAS gestalt reported in the literature. Some disagreement in the characteristic FAS facial shape between the results obtained and those reported in the literature may have been due to ethnic variation.

The advantage of using geometric morphometrics in the assessment of the FAS facial phenotype is the ability to give a comprehensive description of the overall facial shape with a small number of landmark measurements that are statistically unrelated (Halazonetis, 2004). The number of biologically homologous landmarks that may precisely and reliably be identified using the stereo-photogrammetric approach, is limited. Biologically corresponding points on soft tissue surfaces such as the cheek, chin and forehead are not easily located, although some shape information found on these surfaces may be useful in characterising the facial phenotypes of some syndromes (Hammond *et al.*, 2004). Another limitation to the approach used in the Mutsvangwa & Douglas (2007) study was that corresponding landmarks had to be chosen from two separate images, contributing to some measurement error while also being a very time-consuming endeavour.

To eliminate the time-consuming process of manually selecting landmarks from stereo images, Grobbelaar & Douglas (2007) introduced an algorithm that automatically found matched feature points on the second of a pair of stereo images, albeit, after manual selection of landmarks on the first image. Eye measurements, namely palpebral fissure length, interpupillary distance, inner-canthal distance and outer-canthal distance, as well as distances that can be used to approximate the circularity of the upper lip, were obtained using the manual method of marking both images, and the method of automatic marking of the second image. Comparison revealed mean differences less than 1 mm between the two methods. The authors reported that the main source of error in the selection of relevant feature points from stereo photographs for facial feature measurement is the selection of the same point at different depths in the two images.

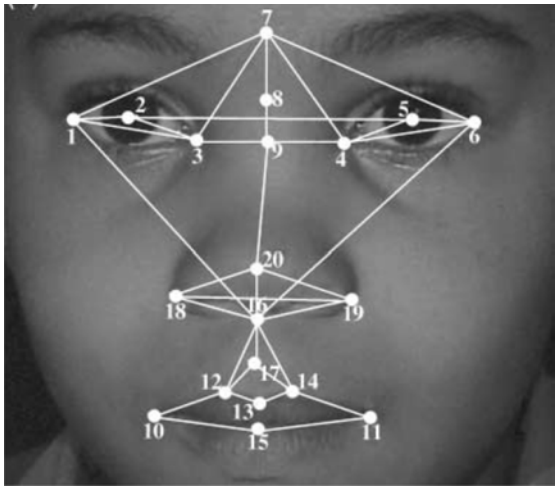

Figure 2.3: Landmarks used in comparing the facial shapes of children with FAS and normal controls: 1, left outer canthion; 2, left pupil centre; 3, left inner canthus; 4, right inner canthus; 5, right pupil centre; 6, right outer canthus; 7, glabella; 8, nasion; 9, sellion; 10, left cheilion; 11, right cheilion; 12, left crista philtre; 13, labiale superius; 14, right crista philtre; 15, stomion; 16, subnasale; 17, midpoint of philtrum furrow; 18, left alare; 19, right alare; 20, pronasale (from Mutsvangwa and Douglas (2007))

2.3.3 Indirect anthropometry: 3D surface imaging

Sophisticated 3D surface imaging modalities have become common in modern indirect anthropometry (Chong & Mathieu, 2006; D'Apuzzo, 2002; Fang *et al.*, 2008; Ghoddousi *et al.*, 2007; Gwilliam *et al.*, 2006; Hajeer *et al.*, 2002; Hammond *et al.*, 2004; Hennessy *et al.*, 2002, 2005; Huang *et al.*, 2005; Kovacs *et al.*, 2006; Majid *et al.*, 2005; Moore *et al.*, 2007; Weinberg *et al.*, 2004, 2006). These optical devices typically use an emitter to illuminate the subject with a light pattern and a sensor system to image the illuminated subject. They offer all the benefits of photography and photogrammetry. The primary advantage is that they provide 3D digital surface images that can be manipulated in 3D space, making measurements easier than using photography-based methods. While photogrammetric measurements cannot be made in areas such as the surfaces on a forehead, cheeks and chin because they are relatively smooth features for which corresponding landmarks cannot be obtained, 3D surface scans allow analysis of the corresponding surfaces. In addition, the problems with matching corresponding landmarks in two or more 2D images are eliminated, improving on accuracy in landmarking (see Figure 2.4).

Examples of 3D surface scanners include structured-light 3D scanning and laser scanning. In structured-light scanning, a pattern projector illuminates the object to be imaged

with some form of coded or structured light. A sensor then images the distorted pattern resulting from reflection by the object and triangulation methods are used to obtain the 3D object data. A colour texture image of the object can be taken rapidly after image acquisition (to prevent motion artefacts) and draped over the 3D image to provide a textured, realistic-looking representation of the subject. Typical accuracy of structured-light systems is approximately 0.5 mm or better (Hennessy *et al.*, 2005). Laser scanners typically use a narrow laser beam that is swept across the subject's face. Digital cameras are used to monitor the sweeping and triangulation geometry is used to calculate depth information.

The major drawback of 3D scanning is cost. Typically, the equipment used is expensive in terms of both hardware and software. Previously, capture speed was an issue, making 3D scanning inappropriate for use on infants and subjects with neurological conditions (Weinberg & Kolar, 2005). Time of capture has improved significantly over the past decade - typically 0.6s for laser scanners (May *et al.*, 2008; Moore *et al.*, 2007) and virtually instantaneous for structured-light photogrammetry. Areas of the head that do not reflect light well, like hair or beards, may cause "holes" in the 3D surface scans. Particular to laser scanners, is that special care has to be taken when dealing with darker coloured skin, which might poorly reflect the projected light from the laser (Majid *et al.*, 2005). In most instances, photogrammetric systems are more modular than laser systems, meaning that they can capture data simultaneously using different devices and integrate the resultant sub-images into a composite image with greater ease (Weinberg & Kolar, 2005).

2.3.3.1 3D surface imaging in FAS diagnosis

Recently, laser scanners have been used in analysing FAS facial anomalies (Fang *et al.*, 2008; Moore *et al.*, 2007). Moore *et al.* (2007) refer to their method as "computerized anthropometry"; they used images obtained with a laser scanner to distinguish between FAS and control subjects with different ethnicities. A total of 276 subjects from four different ethnicities were included in the study. Eighteen percent were North American Caucasian, 9% were African American, 36% were Finnish Caucasion and 37% were of mixed ancestry from the Cape region in South Africa. Sixteen facial measurements were obtained from the scanned images of the subjects. Discriminant function analysis was used to identify the combination of age and facial measurements which best classified subjects into FAS-affected or controls. They reported that ocular-related measurements were included in

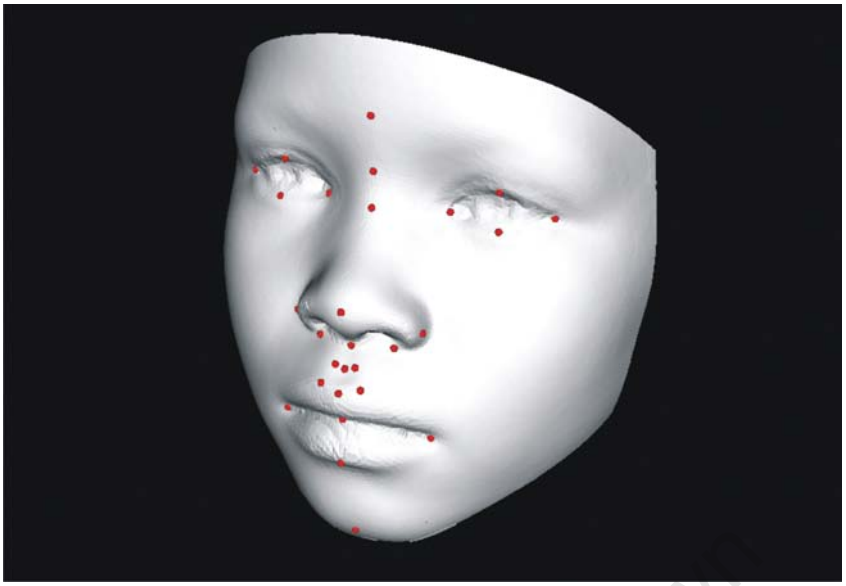

Figure 2.4: A surface scan: The landmarking is much easier on a 3D composite scan than in the 2D photogrammetry case where corresponding landmarks need to be identified in two or more images. In addition, surface information is now available for analysis.

the definition of the FAS facial phenotype in all the ethnicities they assessed. Eight facial measurements identified the Finnish Caucasian group with 93% accuracy; five craniofacial measurements and age correctly classified 92% of the mixed ancestry ethnic group from the Cape region of South Africa; two craniofacial variables correctly classified the African American group with 79% accuracy; 77% accuracy was achieved for the North American Caucasian group using two ocular variables. Their findings were consistent with the research opinion that ethnic variations occur in facial anthropometry (Douglas & Viljoen, 2006; Hall *et al.*, 1989). Fang *et al.* (2008) automatically detected facial features using 3D facial images from a laser scanner. Computer graphics, machine learning and pattern recognition methods were employed to determine a list of features that best differentiate subjects with FAS and normal individuals. Their method, which was applied to two different ethnic populations (50 FAS and 32 control Finish Caucasian; 36 FAS and 31 control mixed ancestry subjects from the Cape region in South Africa) could correctly classify: the Finish sample with 88.2% sensitivity and 100% specificity, the mixed ancestry sample with 91.7% sensitivity and 90% specificity; and the combined group with 82.75% sensitivity and 76.2% specificity. They did not report on the visual differences in facial features between FAS and control subjects in their analysis. Rather, their methodology selected different facial features depending on which pattern classification technique they

used. Their findings confirmed findings by Moore *et al.* (2007) that there exist ethnic differences in the FAS facial phenotype.

2.4 FAS diagnostic criteria

Several diagnostic guidelines for clinical use have been published for FAS. Astley (2006) reports on the five different diagnostic guidelines that have been adopted by different research groups over the previous 10 years. Table 2.5 is an adaptation from that report, showing the two most popular of five diagnostic guidelines used, namely, the 4-Digit Diagnostic Code (Astley & Clarren, 2000) and the Hoyme Diagnostic Guidelines (Hoyme *et al.*, 2005). In common across all five criterion guidelines is the presentation of the patient with growth retardation, some specific facial anomalies, central nervous system developmental problems and a confirmed prenatal exposure to alcohol. They differ however, in terms of the subsets of facial anomalies to include, the extent and manifestation of CNS abnormalities, the level of growth retardation and whether or not knowledge of maternal drinking limits the diagnosis. This illustrates one of the limitations of the current diagnosis paradigm, which is the lack of uniformity in the research community in delineating the FAS facial phenotype.

	4-Digit Code (2004)	Hoyme (2005)
Growth	Prenatal or postnatal height or weight of ≤10th percentile (growth ranks 2, 3, or 4)	Prenatal or postnatal height or weight of ≤10th percentile
Face	All 3 of the following (facial rank 4): PFL of ≤3rd percentile, philtrum rank 4 or 5, and lip rank 4 or 5	Two of the following 3: PFL of ≤10th percentile, philtrum rank 4 or 5 and lip rank 4 or 5
CNS	At least 1 of the following (brain rank 3 or 4): occipital frontal circumference of ≤3rd percentile (microcephaly), abnormal structure, seizure disorder, hard signs, ≥3 domains with impairment ≥2 SDs below the mean (domains may include but are not limited to cognition, memory, language, executive functioning, and attention-deficit/hyperactivity disorder), or global deficits	At least 1 of the following : OFC ≤10th or abnormal structure
Alcohol	Confirmed (alcohol rank 3 or 4) or unknown (alcohol rank 2)	Confirmed to be excessive or unknown

Figure 2.5: Two diagnostic guidelines commonly used in FAS diagnosis.

Chapter 3

Design of a stereo-photogrammetry imaging tool

3.1 Introduction

This chapter deals with the development of a stereo-photogrammetric imaging tool. It begins with an outline of the theoretical background of photogrammetry. The central ideas in photogrammetry, namely perspective projection, interior and exterior orientation parameters, bundle adjustment and calibration are described. The design and construction of the stereo-photogrammetric imaging tool with emphasis on the individual components of the system are described. Calibration procedures for the cameras and calibration frame are presented together with some tests on the accuracy of the procedures. The chapter finishes off with a discussion of the accuracy test results and the system's preliminary performance.

3.2 Theoretical background

Photogrammetry encompasses methods of image measurement and interpretation to derive the shape and location of an object from one or more photographs of that object (Luhmann *et al.*, 2006). The primary objective of photogrammetry is to establish accurately the geometric relationship between the image and the object imaged thereby making possible an accurate 3D reconstruction from 2D images of the object in digital form (coordinates and derived geometric elements) or graphical from (images, drawings,

maps) (Luhmann *et al.*, 2006; Mikhail *et al.*, 2001). The shift from 3D object space to 2D images results in information loss due to (Luhmann *et al.*, 2006):

- Loss of visibility. Anything not in the image cannot be reconstructed back to 3D, for example, the rear of an object.
- Geometric changes. These can be caused by camera positioning relative to the object, camera lens distortion and perspective imaging.
- Colour changes. These may be caused by the effects of the transmission medium (air, glass) on the electromagnetic radiation carrying information.

To accurately reproduce a 3D scene from 2D images there is a need to record the geometric and environmental conditions existing at the time of imaging. The geometric basis of photogrammetry is perspective projection. Mikhail *et al.* (2001) define perspective projection as the transformation of data from a higher dimensional (3D for the object physical space) to a lower dimensional space (2D image). The principles describing perspective projection can be modelled using an idealized pinhole camera (see Figure 3.1). Rays of light reflect from an object and enter through an infinitesimally small aperture. The intersection of the light rays with the image plane produces the image of the object. Thus there exists a collinearity between the object, the pinhole and the image point as illustrated in the pinhole figure.

The interior orientation model of a camera defines camera characteristics needed to reconstruct the bundle of rays with respect to the object space from the corresponding image points (Mikhail *et al.*, 2001). The characteristics include: 1) the perspective centre O which is the reference point where all the light rays pass through; 2) the principal distance or focal distance c is the distance between the image plane and the perspective centre; 3) the position of the principal point relative to the reference system fixed within the image plane (here represented by x'_0, y'_0) and 4) a description of the lens distortion (here represented by Δx, Δy) which may cause a departure from the ideal central projection. The pinhole camera has unlimited depth of field but needs a very long exposure time to create the image. Introduction of a camera lens maintains a sharp image if the pinhole is to be enlarged to reduce exposure time, although a camera lens may lead to distortions. Lens distortions can be divided into two types, namely decentring and radial distortions. Radial distortion represents a radial displacement of the imaged point from its position without distortion from the lens. Decentring distortion refers to the tangential displacement of the image from its correct position (Mikhail *et al.*, 2001). These distortions need to be

accounted for when performing camera calibration by formulating error corrections to the observed image coordinates. The exterior orientation model establishes the position and orientation of the bundle of light rays with respect to the object space coordinate system. For each bundle of rays, six independent elements are needed to describe the exterior orientation, three for position (X_0, Y_0, Z_0) and three for orientation (ω, φ, κ). Fixing the three elements of position involves setting the projection centre for the bundle of rays. The three orientation elements can then be used to orientate the bundle of rays uniquely.

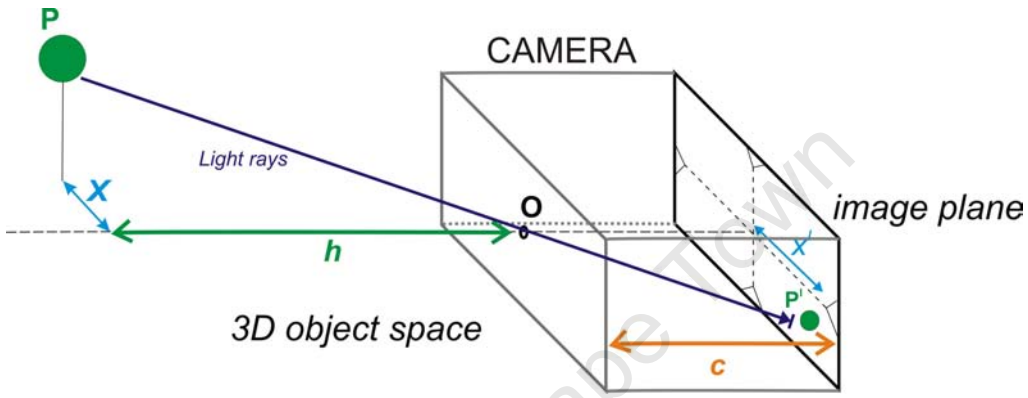

Figure 3.1: The idealized pinhole camera model where P is the object in 3D object space and P' is the projection of P onto the image plane; h is the object to camera distance and represents depth; O is the perspective centre; X is any object distance in 3D object space and x' is the corresponding distance in the image plane; finally, c is the principal distance or focal length.

The equations linking the interior and exterior orientations are called the collinearity equations and form the fundamental equations of photogrammetry. The collinearity equations are below (for a derivation from the collinearity condition see Luhmann *et al.* (2006)):

$$\begin{aligned} x' &= x'_0 + z'\frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} + \Delta x' \\ y' &= y'_0 + z'\frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} + \Delta y' \end{aligned} \tag{3.1}$$

These equations describe the transformation of object coordinates (X, Y, Z) into corresponding image coordinates (x'_0, y'_0) as functions of the interior orientation parameters

(x'_0, y'_0, c, Δx, Δy) and exterior orientation parameters (X_0, Y_0, Z_0, ω , φ, κ) (Luhmann *et al.*, 2006). The three independent rotations ω , φ, κ are represented here by rotation matrix elements r_{11} to r_{33}. The image scale is the deciding factor in the resolution of images as well as in measurement accuracy. Any measurement error in the image is multiplied by the scale factor in the object space. Image scale m, can be described by the relationship between object distance h and the principal distance c or as the relationship between a distance in object space X and the corresponding distance in image space x' (Luhmann *et al.*, 2006):

$$m = \frac{h}{c} = \frac{X}{x'} \tag{3.2}$$

There is a one-to-one mapping of every point in object space to a point in the image plane but the reverse is not true, i.e. a point in the image plane does not map to a point in object space but to a line. Triangulation is the principle used in photogrammetry to produce 3D point measurements. Use of multiple cameras leads to line convergences in image space that uniquely define a point in object space. The determination of interior orientation parameters in photogrammetry is defined as calibration. The imaging environment and the mechanical integrity (e.g. quality of lens) of the cameras influence the stability of the internal orientation parameters. Accuracy specifications determine the effort one needs to go to in modelling the interior orientation parameters; thus, depending on the level of precision and reliability of measurement one would like to achieve, certain assumptions can be made about how to treat the internal orientation parameters and therefore what kind of calibration one may employ (Luhmann *et al.*, 2006). The determination of the exterior orientation of an image is called space resection. Direct measurement of the exterior orientation is not normally possible and thus indirect methods are often employed which exploit the collinearity equations directly or are based on projective relations. Two methods are now discussed. The first, called the direct linear transform, calculates the exterior orientation using projective relations. The other, called bundle adjustment, can be used to simultaneously calibrate camera parameters, orientate multiple images, estimate 3D object coordinates (multi-image triangulation), and model any additional parameters, while providing statistical information about the accuracy and reliability of the resultant model (Luhmann *et al.*, 2006).

3.2.1 Direct Linear Transform

The direct linear transform (DLT) (Abdel-Aziz & Karara, 1971) is a simple and fast linear camera calibration method which models the transformation between the image coordinate system and the object space coordinate system (Mikhail *et al.*, 2001). The DLT equations can be related easily to the collinearity equations ((3.1))(for the derivation of the equations see Mikhail *et al.* (2001)) and are shown below:

$$x = \frac{L_1(X) + L_2(Y) + L_3(Z) + L_4}{L_9(X) + L_{10}(Y) + L_{11}(Z) + 1}$$
$$y = \frac{L_5(X) + L_6(Y) + L_7(Z) + L_8}{L_9(X) + L_{10}(Y) + L_{11}(Z) + 1} \tag{3.3}$$

where x and y form the image space coordinate system, $[X, Y, Z]$ form the 3D object space coordinate system and L_1 to L_{11} are the DLT parameters. Solving these parameters using least squares adjustment determines the relationship between object space and image plane reference systems. A minimum number of six control points with known 3D coordinates is required to solve the DLT in its original form. The control points must not be coplanar, so that they produce a control volume in the 3D object space. The DLT method is popular because of its ease of use and because no initial approximations to the parameters are required. The lens error introduced to the system can be modelled by additional terms but generally the method suffers from simplified camera modelling (the method cannot be used to model image coordinate and reference point coordinate errors) leading to low accuracy results (Luhmann *et al.*, 2006; Remondino & Fraser, 2006) and is not as rigorous as bundle adjustment.

3.2.2 Bundle adjustment

Bundle adjustment or bundle block adjustment is a method for the simultaneous numerical fit of an unlimited number of spatially distributed images (Luhmann *et al.*, 2006). The bundle adjustment model is based on the non-linear collinearity equations (Mikhail *et al.*, 2001; Remondino & Fraser, 2006). Bundle adjustment can use either no control information via inner constraints or, to obtain a better solution, some control marker information. A more advanced bundle adjustment includes calculation of the interior orientation parameters simultaneously in a process called self-calibration (Luhmann *et al.*, 2006). The main advantages of bundle adjustment are that it does not require calibration,

it has a high degree of freedom, thus it increases the accuracy of the solution, and it is considered the most accurate and flexible method of triangulation. It suffers from high computational requirements and the need for good initial camera orientation estimates to guarantee convergence.

3.2.3 Practical considerations

Some practical considerations need to be taken into account when designing a photogrammetry imaging solution. Knowledge of the kind of data one would like to measure is paramount and determines the imaging equipment to be used, the accuracy level to be attained and the configuration of the imaging network. Imaging equipment features affecting accuracy include camera resolution, camera internal stability and type of lens. The lenses, especially, are critical in design as they determine image scale. The desired object coverage determines the number of cameras and configuration of the imaging station. The imaging environment may also become critical in choosing a solution. Lighting conditions determine whether or not flash photography will be necessary. Availability of electricity on location determines the choice of power supply.

Remondino & Fraser (2006) summarize a list of practical calibration considerations. Accurate calibration requires a network configuration of highly convergent images. Errors in interpolating the depth axis are expected to be greater than in the other directions. The distance of separation of the cameras determines the relative resolution of depth in relation to that of the other directions.

Control points are necessary in the DLT and may be beneficial in bundle block adjustment. Bundle adjustment however is more dependent on the number and configuration of imaging devices than on control points (Luhmann *et al.*, 2006). Control points are typically reflective white material in the form of circles or spheres that should ideally be placed on dark, typically, black, backgrounds for maximum contrast. Theoretically, the more control points, and the more widely distributed they are through the 3D volume to be calibrated, the more accurate the calibration. Good control point placement should have control points filling the complete image, or at the very least the edges of the images should be well represented. Calibration accuracy, however, cannot be improved solely by increasing accuracy of control point position or by using more control points. Orthogonal roll (A 90 degree rotation of the camera) of the images must be present to prevent the projective coupling between the interior orientation parameters and the exterior out-

put parameters. One may use a planar instead of a 3D array of controls points in the calibration provided the orthogonal roll condition is adhered to.

The next sections report on the design, construction and calibration of a multi-image (n=3) stereo-photogrammetric imaging system for use in FAS diagnosis using the principles outlined above.

3.3 The stereo-photogrammetry imaging tool

An imaging tool was required which would be capable of imaging a large age range from infants to adults. The system would consist of a calibration frame around the subject's head and three cameras that would be triggered simultaneously to obtain images of the subject's head from three slightly different views using off-the-shelf high resolution digital cameras. Three-dimensional landmark coordinates would then be easily extractable from the 2D images using photogrammetry.

3.3.1 Design specifications

The aim was to develop an imaging tool that would:

1. Provide a low cost solution to the imaging task at hand (obtaining landmark-based facial information) while maintaining an acceptable degree of accuracy for clinical anthropometric evaluations.

2. Be able to provide high resolution images of subjects' full faces from three perspectives and thus provide a good coverage of the face.

3. Be portable for use in field work. The primary use of the instrument would be in FAS screening initiatives in remote areas of South Africa. The system therefore had to be light, portable, and easy to assemble.

4. Be simple to use so that an operator could be trained to use it reliably with sufficient accuracy.

5. Be able to provide a large number of images of subjects between changes of digital memory.

6. Be able to image a large age range, from infants to adults.

7. Be able to be calibrated in the field to monitor internal orientation stability and therefore maintain accuracy of the 2D to 3D mapping.

3.3.2 Calibration frame

The calibration frame structure consists of a backbone onto which the control frame is clamped. The whole structure is painted in matt black for maximum contrast with the control markers. The calibration frame has 91 retro-reflective markers (3M Scotchlite Minnesota, USA), 5 mm in diameter (see Figure 3.2).

Figure 3.2: The calibration frame structure (i.e. a backbone frame and control frame) has 91 retro-reflective markers.

The control frame may be adjusted to facilitate easier placing of the subject. It then swivels into place and comes to rest with vertical struts on either side of the subject's head, defining a 3D volume of control markers. In this position the control frame is hidden from the peripheral vision of the subject so as not to irritate them. The relative configuration of the control markers is maintained by clamping the swivelling component of the frame after it comes to rest around the subject's head. The control frame can also be adjusted vertically to accommodate subjects of different head size and height although this adjustment changes the relative positions of the control markers, requiring the control frame to be recalibrated (see Figure 3.3). To measure younger subjects, a car seat can also be clamped onto the backbone frame for placement of infants. The car seat can accommodate subjects from an age of two months to 5 years.

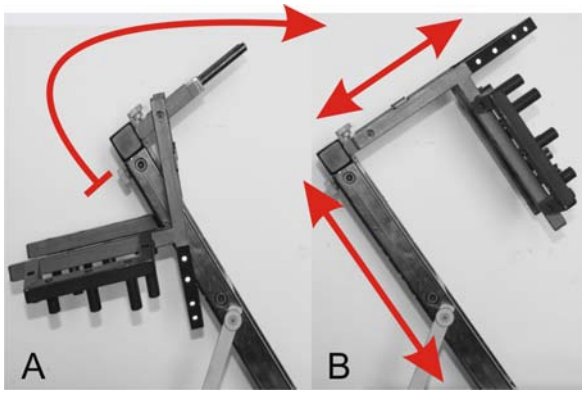

Figure 3.3: The degrees of movement of the control frame relative to the backbone frame as viewed from the side. The control frame can be swiveled out of the way for easier subject placement as in A, and also adjusted for height and head size as in B.

3.3.3 Camera base and imaging components

The rest of the imaging station consists of a camera base clamped onto a tripod. The camera base is made from two aluminum horizontal platforms joined together by vertical struts to maximize torsional rigidity. The tripod can be adjusted for height and angle toward the control frame to accommodate various imaging configurations. Each camera can be securely clamped in position between the horizontal platforms during imaging and removed for storage. Three high-resolution digital cameras are used for image capture (Canon 350D Digital SLR with 18-55mm variable-lens). Each produces an 8 mega-pixel image corresponding to a maximum on-screen resolution of 3450 (width) by 2304 (height) pixels (0.006×0.006 mm). The three cameras are hard-wired to a purpose-built single remote switch that triggers all three simultaneously. All cameras are equipped with DC adapter kits (Canon ACK-DC20) so they are continuously online. The images are stored on Sandisk Extreme 1V 4 gigabyte flash memory cards (Sandisk Corporation, Milpitas, California, USA). Figure 3.4 shows the complete imaging system configured to image an infant. The typical object to image plane distance is 1m although the system can accommodate any distance as long as the subject remains in focus. Should a change in focus length become necessary because the photographic environment requires it, the cameras need to be re-calibrated as described later in this chapter.

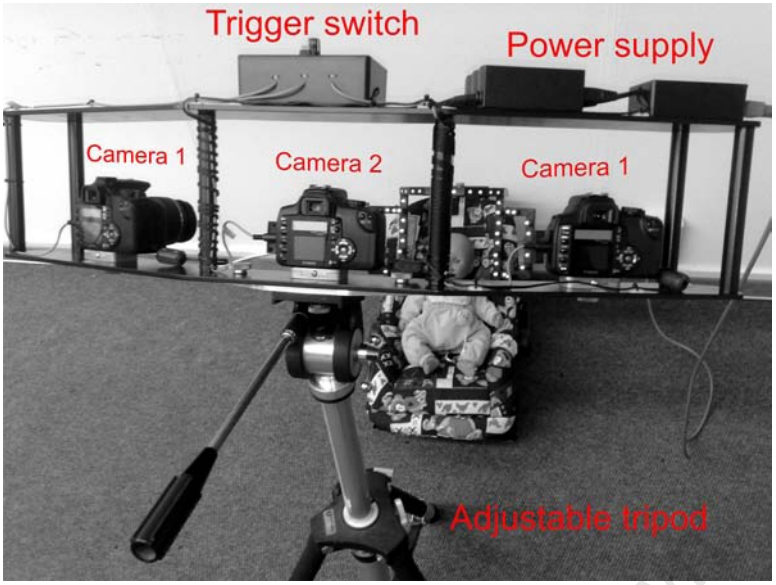

Figure 3.4: The camera base with cameras, trigger and power supply clamped in place. In the background is the calibration frame configuration for imaging an infant.

3.3.4 The remote switch

A purpose-built remote switch was developed to trigger the cameras. The remote switch is hard-wired to all three cameras through a microprocessor. Three wire leads with stereo-jacks link the microprocessor output to the cameras while simultaneously ensuring the cameras are electrically isolated from each other. The input to the microprocessor is a single reset switch which in turn sends three simultaneous pulses to the cameras. Pressing the reset switch prepares the cameras to take images and releasing it releases the shutters.

3.3.5 Software used in analysis

The software used in all the 3D object measurements was Australis software Version 6.06 (Photometrix Pty. Ltd., Victoria, Australia). This software can perform automated off-line high-precision metrology measurements from convergent digital image networks, either using digital cameras or scanned film imagery. An integrated image measurement, preliminary orientation and bundle adjustment approach can obtain 3D object point co-ordinates and camera calibration data from multi-camera, multi-image networks of an effectively unlimited number of object points (Photometrix, 2004).

3.4 Calibration and testing of tool

3.4.1 Determining 3D coordinates of the calibration frame

The 3D coordinates of the calibration frame have to be determined with a high degree of accuracy if they are to function as 1) 3D reference points for bundle adjustment during object coordinate measurements and 2) as a calibration tool for cameras in the field. To determine the relative 3D coordinates of the centres of the retro-reflective markers, stereo-photogrammetry was employed. The calibration frame was placed at the centre of a laboratory grid containing retro-reflective markers. The 3D coordinates of the centres of the circular laboratory grid markers were known to a high degree of accuracy (0.1mm). A high resolution camera was used to obtain images of the laboratory grid from various positions to obtain a set of sixteen highly convergent images (see Figure 3.5). During imaging, the camera was held at a constant focal point and the lens was taped down to maintain internal orientation stability. The calibration frame reflective marker centres were then selected by mouse click in each image. The Australis software automatically detected the centre of any circular marker selected on screen. Bundle adjustment was used to determine their 3D coordinates to a high degree of accuracy.

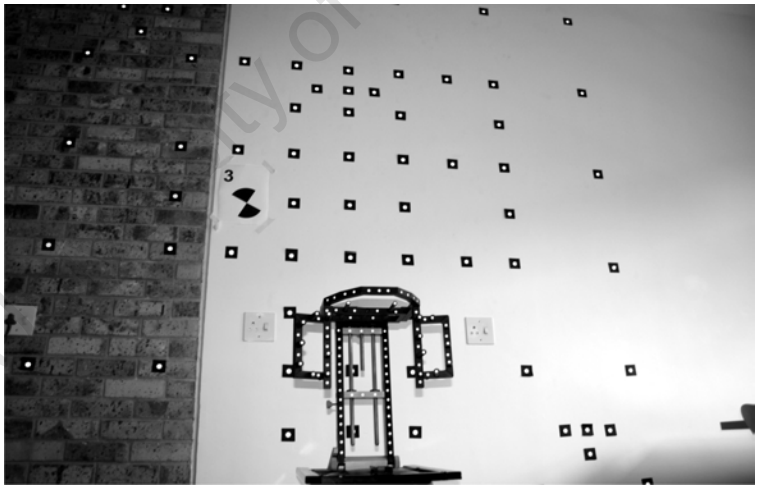

Figure 3.5: The calibration frame situated in the middle of a laboratory calibration grid. The accurately known 3D coordinates of the markers in the laboratory calibration provide reference points. Sixteen photographs with roll diversity were taken from various angles and distances to produce the highly convergent network necessary for bundle adjustment.

The control frame component of the calibration frame contains 47 retro-reflective markers (see Figure 3.6). During image acquisition, only the control frame markers are used. All retro-reflective markers (91), however, are used in field camera calibration. To check the accuracy of the 3D coordinate calibration, the control markers (n= 47) were used to determine the 3D coordinates of the check markers. See Figure 3.6 for differences between control and check markers. From the set of check markers, (n= 23) were chosen because their position would coincide with the volume of best interpolation for the imaging control markers and they would afford insight into the interpolation capabilities of the control frame. Choo & Oxland (2003) report that control points can more accurately interpolate than extrapolate the 3D volume surrounding them. The subject's head is situated in this volume of best interpolation during imaging. The control markers function in the same way as the laboratory calibration grid described above in that they can be used as reference control points to calibrate the check markers. The comparison of the known 3D coordinates (found accurately using the laboratory calibration grid) and those determined during the interpolation assessment was used as an indicator of accuracy. The results of the control frame interpolation assessment showed that the known 3D coordinates (those found using the laboratory calibration grid) of the check markers and those determined during the test were comparable. The differences in the X and Y axes were small with a maximum value of 0.31 mm in the X direction. The Z or depth coordinate errors however, ranged from 0.11 mm to 1.29 mm with mean, median and standard deviation values of 0.56 mm, 0.40 mm and 0.37 mm, respectively. An extension of the assessment to include the remaining 21 check markers revealed errors in the Z axis of up to 2 mm in magnitude. The check markers furthest from the area of best interpolation of the control frame showed the largest differences between measured and known coordinates particularly in the Z direction.

3.4.2 Camera calibration using the control markers

A typical camera calibration procedure using the calibration frame involves fixing each camera's focal length at an object-to-image distance of approximately 1m, with the calibration frame filling the image. Each camera is set to manual focus and clamped using a purpose-built vinyl clamp to prevent camera movements from affecting the lens stability.

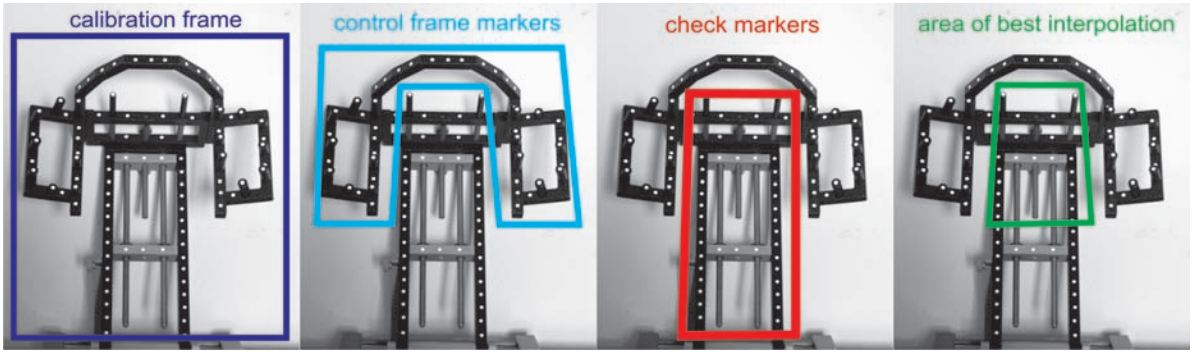

Figure 3.6: The calibration frame showing the different constituent regions. There are 91, 47, 44 and 23 markers in the calibration frame, control frame, check marker and area of best interpolation, regions, respectively.

A set of sixteen[1] highly convergent images of the calibration frame are obtained with at least four of the images orthogonally rolled. Using the 3D coordinates of the dense array of control markers and self-calibrating software procedures of Australis, the camera's interior orientation parameters are determined. Once the interior parameters of the camera are known they are fixed for any measurements taken from images taken by the camera. This procedure is one that should be carried out before any imaging session, or should the focal length of a camera change, to maintain photogrammetric accuracy. Figure 3.7 shows the control frame and the camera positions during camera calibration as viewed in Australis software.

3.4.3 Camera calibration drift

In the field, camera shaking and other environmental factors may disturb the calibration settings of the cameras. Once the parameters are determined during a calibration session, they must remain fixed throughout all subsequent imaging although it is technically feasible to check the camera calibration for each image, albeit from a single image and weaker control point geometry. An assessment of calibration drift with time was performed. Three calibration sessions were performed within three days without changing any of the camera settings, in order to ascertain if the camera principal distances changed with time and different environmental conditions. The principal distance of focal length

[1]Eight images are usually enough Luhmann *et al.* (2006) but 16 was used to increase redundancy and include more images with roll diversity

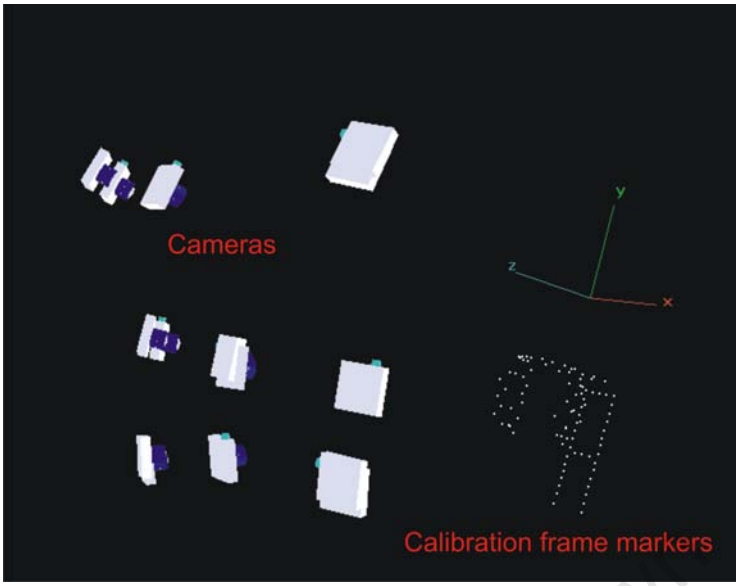

Figure 3.7: Sixteen (some cameras are obstructed from view) highly convergent camera positions of the calibration frame as shown in the Australis software. The relative positions and orientations of the cameras are shown together with the 3D coordinates of the calibration frame after bundle adjustment.

is used here as representative of interior orientation parameters. The results of three sessions of camera calibration are shown in Table 3.1. The object to camera distance was similar and all cameras had the same settings during calibration resulting in similar principal distances.

Table 3.1: Calibration drift results: principle distances obtained after three days.

	Camera 1 (mm)	Camera 2 (mm)	Camera 3 (mm)
Session 1	49.660	49.270	49.510
Session 2	49.680	49.260	49.480
Session 3	49.640	49.260	49.440
Mean	49.660	49.263	49.477
Std	0.020	0.006	0.040

3.4.4 Camera synchronization

The reliability of the synchronization of the trigger switch was assessed using an oscilloscope trace. The time/division was set at 0.5 ms on the oscilloscope. It therefore took 5 ms for a waveform to be swept across the screen. Any instantaneous differences in waveform position between images taken simultaneously could thus be calculated. Figure 3.8 shows one such test with a waveform imaged by each camera as it sweeps from left to right on the oscilloscope. The shutter speeds for the cameras were set at f/2000 to prevent the wave trace from cycling more than once during exposure. The calibrated cameras were then positioned to converge on the oscilloscope screen and triggered. The images were then compared for waveform position and therfore time differences. A set of ten tests were performed for camera synchronization errors. The results are shown in Table 3.2.

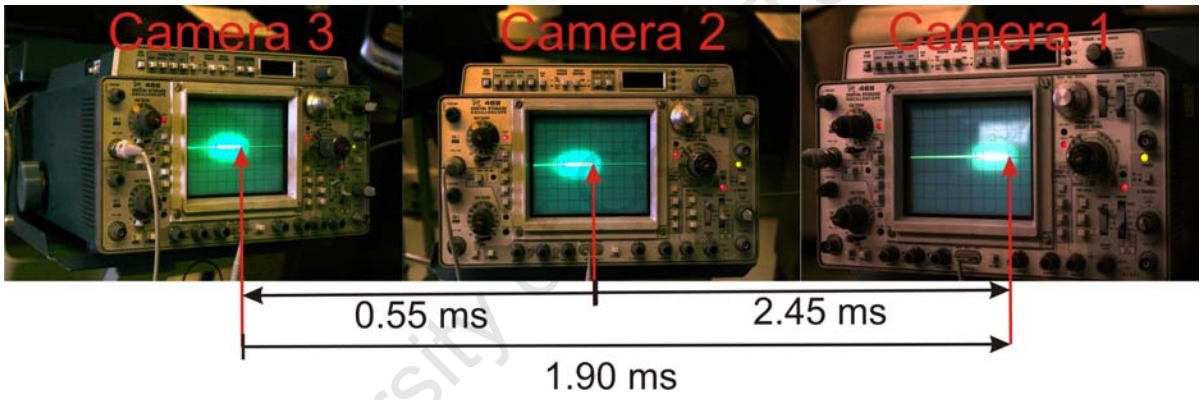

Figure 3.8: Synchronization using an oscilloscope. Three different waveform positions from the three cameras triggered simultaneously. The figures indicated show the time differences as determined by the waveform positions, during Trial 4 (Table 3.2).

3.5 Chapter discussion and original contribution

Close range stereo-photogrammetry has become popular in craniofacial landmark anthropometry for growth studies, genetics, and surgical planning (Chong & Mathieu, 2006; Majid *et al.*, 2005; Ras *et al.*, 1996b) and as presented here in syndrome diagnosis (Meintjes *et al.*, 2002). It easily lends itself to use in medical imaging because it is non-invasive and image acquisition is virtually instantaneous. Commercially available systems such as

Table 3.2: Synchronization results: Absolute differences in ms

	Camera 1 and 2 (ms)	Camera 2 and 3 (ms)	Camera 3 and 1 (ms)
Trial 1	0.80	0.75	1.55
Trial 2	0.60	0.30	0.30
Trial 3	0.95	0.90	0.05[a]
Trial 4	0.55	2.45[b]	1.90
Trial 5	0.80	1.20	0.40
Trial 6	0.15	1.30	1.15
Trial 7	0.50	0.05	0.45
Trial 8	0.15	0.50	0.65
Trial 9	0.45	0.10	0.35
Trial 10	0.10	0.05[a]	0.05[1]
Average	0.51	0.76	0.69
Std	0.30	0.75	0.64

[a]lowest difference
[b]highest difference

the Vectra 3D (see Section 4.5) offer comparable performance, in terms of image acquisition times and portability, to the system presented here (Hammond *et al.*, 2004; Honrado & Larrabee, 2004). These 3D systems produce virtual 3D surface images that can be rotated for easier landmark identification, which offers a significant advantage over 2D stereo-photogrammetry. They are, however, more costly. Laser scanning is also vulnerable to motion artefacts from subject movement during image capture since most systems require at least two scanner shots to obtain a complete image of the face (Kovacs *et al.*, 2006). The system presented here is useful in large-scale surveillance studies because the equipment is modular, hence easy to transport, and provides low cost per measurement as the components are generally cheap. Aside from the cameras, camera accessories and software, the system cost less than US$ 200 to construct. The prices of high resolution digital cameras are continuously falling. The system image acquisition is not labour intensive and is also easy to learn. Assembling the system takes a relatively short time, typically 15 minutes.

The camera calibration results show that the control frame can be used to calibrate the cameras reliably and consistently. This means that the cameras can be calibrated in the field where they might be located for long periods e.g. during large-scale epidemiological studies. Another advantage is that the same environmental conditions exist between

calibration and imaging. The 3D coordinates of the check markers in the volume of best interpolation were found to high accuracy in the interpolation assessment of the control frame; this procedure highlighted the importance of placement of the subject's head in the area where best interpolation is achieved. Depth errors were larger than those of other dimensions, as was expected. The camera-to-object distance and the positions of the image control markers were designed to minimize error in depth around the volume where the subject's head is positioned. The majority of the check markers are behind the subject's head in object space and since the interpolation was acceptable, it can be concluded that depth errors can only decrease toward the optimized area. The extension of the assessment to include the remaining 21 check markers confirmed the inability of the image control markers to extrapolate accurately outside the control area.

Subjects are prone to movement during image acquisition with muscle flinching and blinking being particularly problematic due to the high speed with which they can occur. High camera shutter speed can be useful in obtaining an image "at the right time" i.e. while the subject is in a specific position or pose. In photogrammetry, however, it is not only important to obtain images while the subject is in the correct pose but it is critical that all cameras capture images at the same instant. It has been reported that the fastest facial movement is a normal eye blink with a velocity of about 206 mms^{-1} (Majid *et al.*, 2005; Somia *et al.*, 2000). At this speed, the largest time delay reported here, namely 2.45 ms, would correspond to a distance of 0.51 mm. Thus the synchronization delays between cameras were considered acceptable for the purposes of facial measurement.

A stereo-photogrammetry imaging device was developed previously for FAS research (see Figure 2.2 in Chapter 2) (Meintjes *et al.*, 2002). The system used two cameras and a control frame similar to the one described here. The direct linear transform (DLT) was used to determine the 3D object space coordinates of facial landmarks. The system described here is an improvement on that system. Three cameras are used instead of two, adding a frontal perspective to the image set and making landmark selection easier and more robust. The limitations of the DLT have been discussed earlier in the chapter; the bundle adjustment used here has the potential to provide a more accurate mapping than the DLT. Previous work using the older stereo-photogrammetry imaging tool yielded encouraging results for inter-landmark distance measurement for FAS diagnosis (Douglas *et al.*, 2003b; Meintjes *et al.*, 2002).

Of special interest is an exploration of geometric morphometrics for use in classifying the FAS facial phenotype in infants, as was previously done for older children using landmarks acquired with the older imaging tool (Mutsvangwa & Douglas, 2007). The

system developed here allows for a a greater age range of subjects, including infants, by including a car seat in which younger subjects can sit while being imaged.

Chapter 4

Stereo-photogrammetric Reliability and Precision

4.1 Background

Strict requirements need to be met in the quantification of form and volume of body structures (Kovacs *et al.*, 2006), thus a thorough assessment of precision and reliability needs to be made for any system designed to take such measurements. Weinberg *et al.* (2004) noted that the ability to obtain precise and accurate measurement data is the principal criterion upon which to evaluate a measurement technology. Various indirect anthropometry systems are available and the degree of correspondence of measurements obtained using these competing modalities has been an active research topic (Aung *et al.*, 1995; Ghoddousi *et al.*, 2007; Kovacs *et al.*, 2006; Weinberg *et al.*, 2004, 2006; Winder *et al.*, 2008). In tests of reliability, the system to be investigated is compared to an established benchmark. In the literature, direct anthropometric measurements have been used extensively as the "gold standard" for assessing reliability in indirect anthropometric methods (Aung *et al.*, 1995; Ghoddousi *et al.*, 2007; Kovacs *et al.*, 2006; Weinberg *et al.*, 2004, 2006). Direct anthropometric measurements for diagnosis of FAS are traditionally done using hand-held rulers, callipers, or a soft measuring tape to obtain linear distances between facial landmarks and protractors to obtain angles (Farkas, 1994).

The methodology of performing direct measurements, whether from a human sample or from realistic-looking models, is now well established (Farkas, 1994). Recently, comparisons between indirect anthropometry modalities have appeared in the literature (Ghoddousi *et al.*, 2007; Weinberg *et al.*, 2006). Weinberg *et al.* (2006) provide motivation

for comparing alternative measurement technologies: 1) the cross-validation of technology provides a way of delineating the idiosyncrasies of a system that can affect the results of a study and 2) it is important to clarify the degree of correspondence obtained using different technologies. The method for assessing reliability is generally the same whether one is comparing direct anthropometry and a new indirect measurement technology, or two indirect measurement technologies. Distance, angles and distance ratio measurements are usually obtained over several sessions and by several operators and then averaged to obtain reference values that are used to assess the measurement technology being investigated. Precision is the level of reproducibility of a measure and imprecision is defined as the within-subject variability caused by inconsistency between repeated measures of the same quantity (Weinberg *et al.*, 2004). Reliability here, as defined for accuracy in Kovacs *et al.* (2006), is the degree of veracity of a measurement and reflects the ability to obtain a measurement that correctly reflects the size of the quantity being measured. Figure 4.1 illustrates the difference between precision and accuracy.

Figure 4.1: The difference between precision and accuracy is illustrated by these three bullseye diagrams. The first shows low accuracy but high precision. The second shows high accuracy but low precision. The last illustrates high precision and accuracy.

This chapter begins by giving a theoretical overview of the statistical metrics used in analysing relaibility and precision. A comparison study of anthropometric measurements obtained from a doll using direct anthropometry and stereo-photogrammetry is then reported, followed by a comparison of the same anthropometric methods using human subjects. Two indirect measurement systems, namely static stereo-photogrammetry and structured-light stereo-photogrammetry, are then compared with respect to measurements on human subjects. A discussion of the performance of the stereo-photogrammetry method concludes the chapter.

4.2 Statistical metrics of precision and reliability

Mean absolute difference (MAD) and relative error of magnitude (REM) are some of the methods used to quantify precision (Jamison & Ward, 1993; Weinberg *et al.*, 2006). Mean absolute difference can be calculated for a measurement by averaging the absolute difference between values at time 1 and time 2 across all subjects in a sample. This precision statistic has the advantage of being easy to calculate, requires few assumptions and is easily interpretable (Weinberg *et al.*, 2006). The relative error of magnitude (REM) in precision is the difference between two measurements as a percentage of the grandmean of the two sets of measurements:

$$REM = \frac{MAD}{grandmean} \times 100 \tag{4.1}$$

where the grandmean is the average of all the measurements in the sample of the same quantity taken at different times. Intra-operator precision is an indication of the consistency with which the same operator measures the same subject at different times, and inter-operator precision is the corresponding indicator of consistency for two operators on the same measurement quantity.

Some practical considerations need to be taken into account when assessing reliability and accuracy. In landmark identification, particularly using indirect anthropometry, the effects of memory influence reproducibility (Weinberg *et al.*, 2006). In that regard, recall bias effects have to be accounted for by allowing for some interval of time to expire between measurements. There is, however, no consensus as to the time period required to minimize recall bias (Gwilliam *et al.*, 2006).

Prior marking of landmarks has been shown to increase precision in anthropometry especially in indirect anthropometry were palpation of landmarks is impossible (Weinberg *et al.*, 2006). This option is available for tests of accuracy on inanimate objects and may reveal the true precision of the system being tested. However, it is not always possible to landmark human subjects, particularly if the subjects are young or syndromic patients. The precision values obtained without prior marking would thus inherently contain error components attributable to incorrect identification of some landmarks in the image. In 3D systems, landmark precision can be calculated along each of the principal axes provided the coordinate system of an image does not change during measurement trials. This has the advantage of delineating errors along each of the principal directions and assessing, for example, the effects of mapping from 2D to 3D.

4.3 Validating the stereo-photogrammetric imaging tool

4.3.1 Methods

The system produces a set of three high resolution colour ($3450(width) \times 2304(height)$ pixels) images in Canon CR2, a Canon RAW format. Image quality can be assessed visually. Images deemed to be poor due to lighting conditions, obstruction and motion artefacts introduced by subject restlessness may be discarded. In this study the images retained were preprocessed in Canon ZoomBrowser EX v5.5 (Canon Information Systems Research Pty. Ltd, Australia) before being converted to monochrome TIFF images. Monochrome images were preferred because they provide higher contrast for control marker identification.

A doll was used to assess the precision and reliability of the stereo-photogrammetric imaging tool. Landmark selection from the digital images was done in Australis software. Corresponding landmarks were selected by mouse click in each of the three images that constitute a scene set. Twenty-four standard facial surface landmarks were pre-labelled on the doll using permanent ink. The landmarks were chosen based on: 1) their relevance to the FAS facial phenotype; 2) the ease of identification using both direct anthropometry and stereo-photogrammetry, and 3) the distribution about the face such that the precision of both short and long inter-landmark distances could be assessed. Figure 4.2 shows the doll with the landmark locations and their descriptions. Care was taken to make the landmark labels as small but circular as possible so that their centre could be sufficiently approximated in the images. To assess the precision and reliability of the system, a comparison of linear distances obtained directly using direct anthropometry and indirectly using the stereo-photogrammetric system was made for all landmarks on the doll. The direct anthropometry measurements were obtained using digital callipers. Measurements taken during two sessions by two operators in each session were averaged to obtain reference values for comparison with those derived from the system. The period between the two sets of measurements was at least 24 hrs to remove recall bias in measurements. The same protocol was observed for measurements using the stereo-photogrammetry imaging tool, namely, two sessions of measurements with measurements obtained from the stereo images by two operators in each session. This resulted in eight sets of 275 inter-landmark measurements (four sets manually obtained and four sets obtained through stereo-photogrammetry). One distance (left tragion to right tragion) could not be used as it proved impossible to measure manually using the callipers.

4.3.1.1 Reliability of the imaging tool

The inter-modality reliability was analysed using tolerance thresholds as defined in Kovacs *et al.* (2006). A stereo-photogrammetry measurement less than 1 mm different from the manual measurement was considered "very reliable"; between 1 and 1.5 mm, "reliable"; between 1.6 and 2 mm, "moderately reliable"; and anything more than 2 mm, "unreliable". In addition, the proportional error measure, REM, was calculated, for each operator and also using the combined scores of both operators. For each operator, measurements from their two sessions for each modality were averaged. These two averages, one from the stereo-photogrammetric measurements and one from the manual measurements were used to calculate the grandmean. The difference between the two sets of measurements divided by the grandmean gave each operator's normalized inter-modality reliability.

The averaged normalized inter-modality reliability was calculated by averaging the two operators' measurements for both the manual measurements (a total of 4 sets of measurements, 2 for each operator) and for the stereo-photogrammetry measurements (a total of 4 sets of measurements, two for each operator). These two sets of averaged values were then averaged to get a grandmean. The difference between the two sets was divided by the grandmean to give a measure of relative error. This adaptation of REM is reported here as the normalized inter-modality reliability to differentiate it from the REM measure used in precision later. Normalized inter-modality reliability scores less than 1% are deemed "excellent"; between 1 and 3.9%, "very good"; between 4 and 6.9%, "good", between 7 and 9,9%, "moderate" and finally anything above 10%, "poor". This thresholding method was adapted from Weinberg *et al.* (2004).

4.3.1.2 Precision of direct anthropometry

Both intra- and inter-operator precision of manually measured inter-landmark distances was assessed by comparing corresponding inter-landmark distances between operators and within operators. For intra-operator precision REM values, time 1 and time 2 distances for each operator were used to calculate the grandmean. The difference in time 1 and time 2 inter-landmark distances was then divided by the grandmean to obtain REM values. For inter-operator precision, two sets of measurements were averaged across each operator. The averages for each operator were then averaged again to obtain the grandmean. REM was therefore the differences between operator averages divided by the grandmean. REM

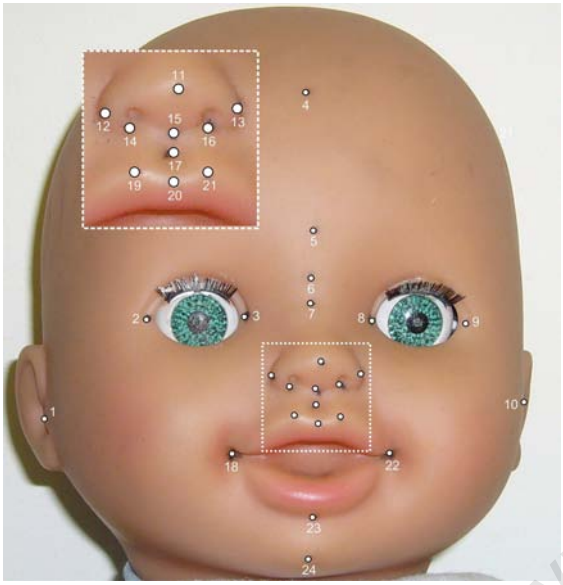

Figure 4.2: The landmarks used in reliability assessment. These landmarks have been identified in the literature as significant in FAS diagnosis. Inter-landmark distances were used to make reliability comparisons between manual (vernier-calipers) and stereo-photogrammetry systems. The landmarks are: 1. tR-right tragion, 2. exR -right exocanthion, 3. enR-right endocanthion, 4. tr-trichion, 5. g-glabella, 6. n-nasion, 7. se-sellion, 8. enL-left endocanthion, 9. exL-left exocanthion, 10. tL-left tragion, 11. psn-pronasale, 12. alR-right alare, 13. alL-left alare, 14. sbalR-right subalare, 15. s-subnasale, 16. sbalL-left subalare, 17. Phl-centre of philtrum furrow, 18. chR-right cheilion, 19. ls'R-right crista philtri, 20. ls-labiale superius, 21. ls'L-left crista philtri, 22. chL-left-cheilion, 23. li-labiale inferius and 24. pg-pogonion.

scores less than 1% are deemed "excellent"; between 1 and 3.9%, "very good"; between 4 and 6.9%, "good", between 7 and 9,9%, "moderate" and finally anything above 10%, "poor".

4.3.1.3 Precision of the imaging tool

The 3D coordinates determined using stereo-photogrammetry and used to calculate inter-landmark distances, were stored during each measurement session by each operator. Precision for stereo-photogrammetry measurements was calculated for each landmark along each of the principal axes. Precision could be calculated for each landmark since the coordinate system of an image does not change during measurement trials. Tolerance thresholds were used in the classification of precision for each landmark (Aldridge *et al.*,

2005): a difference of less than 1 mm between two sets of measurements was considered "highly precise"; between 1 and 1.5 mm, "precise"; from 1.6 to 2 mm, "moderately precise"; and anything more than 2 mm, "less precise". Intra-operator precision was calculated along each axis using the measurements from each operator's first and second session. Inter-operator precision was calculated using the mean of each operator's measurements.

4.3.2 Results

4.3.2.1 Reliability of the imaging tool

The results of the inter-modality reliability are presented in Table 4.1. Of the total of 275 inter-landmark distances, 92.4% of the inter-landmark distances (averaged between operators for each modality) were "highly reliable" in a comparison between stereo-photogrammetry and direct anthropometry. None were "unreliable". The stereo photogrammetric distances were however, generally larger by an average of 0.4 mm with only 8.7% of the direct anthropometry data having larger values. The normalized inter-modality reliability values in Table 4.2 show that 96.0% of both operator 1 and inter-operator values and 97.1% of operator 2 values were at least good. Poor measurements accounted for 0.7% of both operator 1 and inter-operator values and 1.1% of operator 2 values. The inter-landmark distances yielding poor results are reported in Table 4.2 for each operator and between operators.

Table 4.1: Inter-modality reliability

	Difference (mm)	Operator 1 ($\%$)	Operator 2 ($\%$)	Average ($\%$)
Highly reliable	< 1	92.4	91.6	92.4
Reliable	1-1.5	7.6	6.9	7.6
Moderately reliable	1.5-2	0.0	1.5	0.0
Unreliable	> 2	0.0	0.0	0.0

4.3.2.2 Precision of direct anthropometry

Figure 4.3 shows the differences between inter-landmark measurements for the two operators using vernier callipers. From a total of 275 distances, 98.7% were highly precise, 1.8% were precise and 0.4% were moderately precise. None of the differences were greater

Table 4.2: Normalized inter-modality reliability (Results are rounded and may add to slightly less or more than 100%)

	Operator 1 (%)	Operator 2 (%)	Averaged between operators (%)
Excellent	29.1	25.8	28.4
Very good	58.6	62.2	60.4
Good	8.4	9.1	7.3
Moderate	3.3	1.8	3.3
Poor	0.7	1.1	0.6
Distances with poor reliability [a]	sn-alR Phl-sn	sbalR-alR sbalR-sn Phl-sn	sbalR-sn Phl-sn

[a]see Figure 4.2 for label definitions

than 1.6 mm. The REM values shown in Table 4.3 reveal that, for operator 1, 97.8% were good, very good or excellent, and the remaining 2.2% poor. Only 1.5% of the results were moderate and 0.7% poor for operator 2, with the remaining 97.8%, good, very good or excellent. For inter-operator precision, 1.5% of results were moderate and 1.5% were poor with the rest, good, very good or excellent. The inter-landmark distances yielding poor results are reported in the table for each operator and between operators.

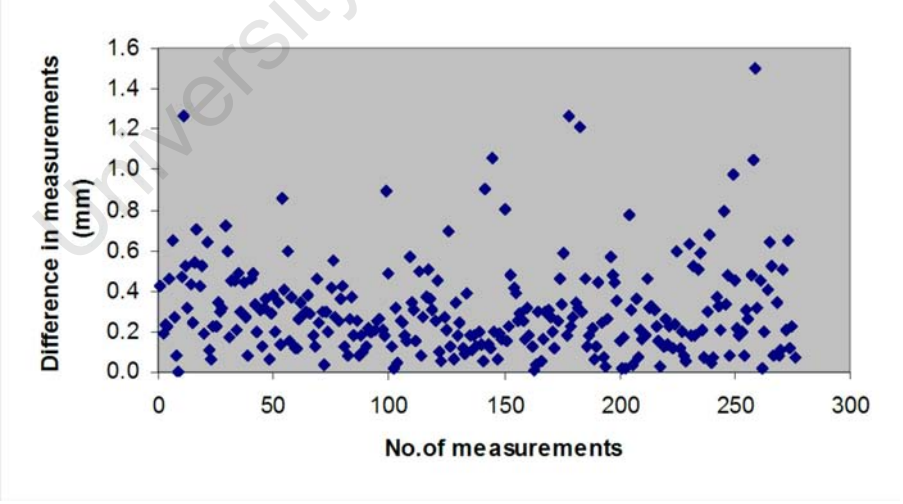

Figure 4.3: Inter-operator differences for direct anthropometry inter-landmark distances.

Table 4.3: Direct anthropometry REM (Results are rounded and may add to slightly less or more than 100%)

	Operator 1 (%)	Operator 2 (%)	Inter-operator (%)
Excellent	53.1	60.4	57.5
Very good	40.7	30.9	35.6
Good	4.0	6.6	4.4
Moderate	0.0	1.5	1.5
Poor	2.2	0.7	1.1
Distances with poor reliability	alR - enR, sbalR - alR, sbalR - sn, Phl - sn, ls'L - Phl, ls'R - ls	sn - alR, Phl-sn	sbalR - sn, ls'L - Phl, ls'R - sbalR

4.3.2.3 Precision of the imaging tool

The stereo-photogrammetry REM precision values for inter-landmark distances are shown in Table 4.4. Operator 1 and inter-operator REM values are all at least good. Only 0.4% of the measurements for operator 2 were moderate, with the rest, at least good. Both intra- and inter-operator precision using landmark coordinates are shown in Figure 4.4. The intra-operator differences for one operator along each axis for each landmark were all less than 0.9 mm. The intra-operator grandmean difference across all axes was 0.1 mm. The values ranged from 0 mm to 0.9 mm with a median of 0.1 mm. The z axis differences were consistently larger than those for the other two axes with the right alare being the least precise measurement on that axis. The z axis median was 0.2 mm. The inter-operator differences along each axis for each landmark were all less than 0.6 mm. The differences in the y axis were less variable than those for the two other axes. Again the z axis differences were consistently larger than those for the other two axes with the labiale inferius being the least precise measurement on all the axes. The inter-operator grandmean difference across all axes was 0.1 mm. The values ranged from 0 mm to 0.5 mm with a median of 0.1 mm; the z axis median was 0.1 mm.

Table 4.4: Stereo-photogrammetry REM (Results are rounded and may add to slightly less or more than 100%)

	Operator 1 (%)	Operator 2 (%)	Inter-operator (%)
Excellent	89.5	85.8	85.9
Very good	9.8	12.4	13.8
Good	0.7	1.5	0.4
Moderate	0.0	0.4	0.0
Poor	0.0	0.0	0.0

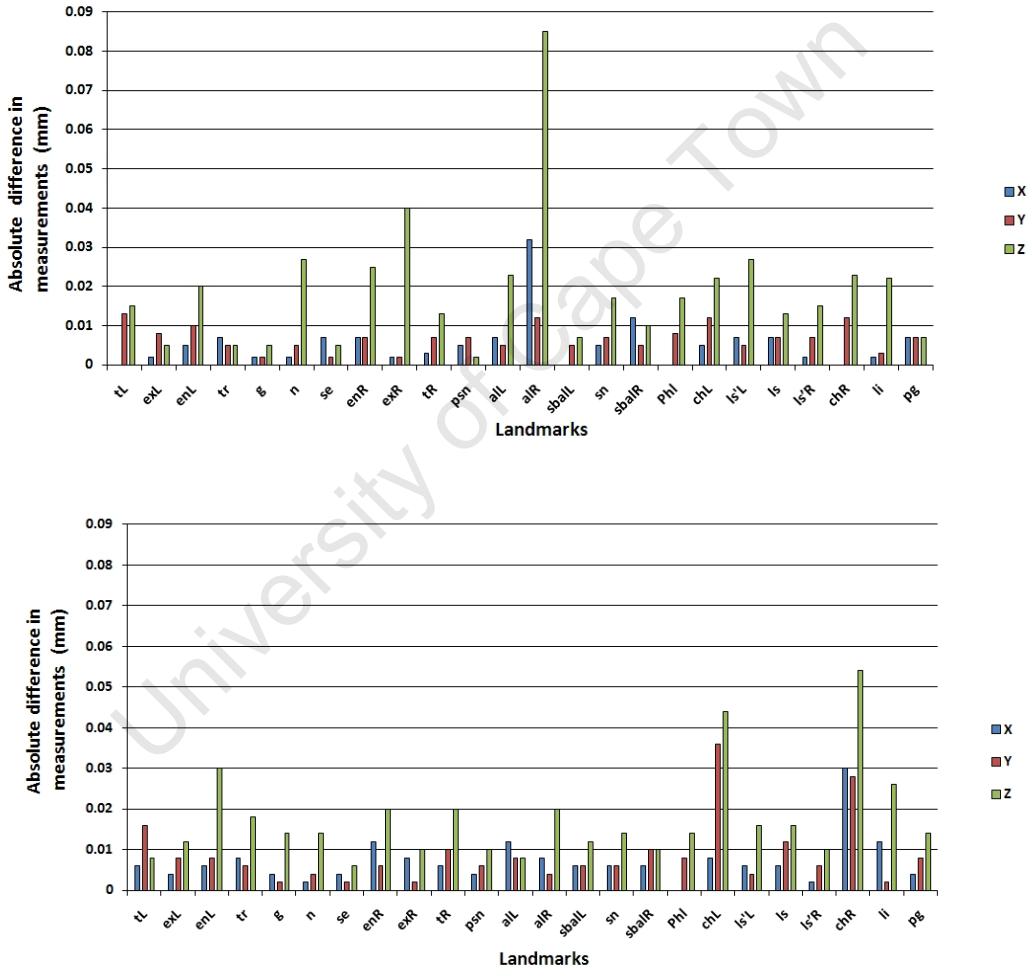

Figure 4.4: Intra- (top) and inter-operator (bottom) precision of the stereo-photogrammetry imaging tool.

4.4 Comparing clinical and imaging tool measurements

A comparison was made between eye measurements taken in a clinical setting by a clinician and those obtained through use of the imaging tool. The study was limited to those eye measurements that would normally be gathered by dysmorphologists during a FAS screening exercise, and no attempt was made to alter the standard procedure, so that the children involved would not be subjected to any additional examination.

4.4.1 Methods

4.4.1.1 Study sample

Study data were obtained during the screening of children for FAS in disadvantaged communities in the Upington and De Aar regions of the Northern Cape Province in South Africa. Thirty healthy subjects were chosen randomly from a total study sample of 254 subjects. The ages of the subjects ranged from 1 to 6 years (mean, 5.4 years; SD, 1.3). Written informed consent for the study was obtained from the children's guardians.

4.4.1.2 Inter-modality reliability of eye measurements

The direct anthropometric appraisal for the study was performed by a single dysmorphologist. The eye distances measured were interpupillary distance (IPD), inner canthal distance (ICD) and palpebral fissure length (PFL) (see Figure 4.5). The clinician took the measurements using a hand-held ruler. For PFL measurements, only one eye was used. The measurements were taken once and recorded. All the measurement values were integers, a limitation introduced by the resolution of the hand held ruler. The imaging was performed after the direct anthropometric appraisal of the subject. One operator then took eye measurements from the resulting images. The statistical measure of reliability described above, namely the normalized inter-modal reliability, was calculated to obtain the proportional error in measurements between the hand-held ruler and the stereo-photogrammetrically derived measurements.

Figure 4.5: Eye distances used to compare measurements obtained manually in a clinical setting to those obtained using the stereo-photogrammetry imaging tool.

4.4.1.3 Estimating the effects of pose

A study to assess the effects that different poses have on the reliability of measurements was performed on five subjects. Since some of the subjects were infants, care was taken to place them in the car seat carefully, in such a way that their heads were surrounded by the calibration frame. Several images were taken until a set was obtained which showed the subject in the correct pose. On average, two sets of images were taken before the operator was satisfied with the pose. Figure 4.6 shows a series of photographs of the same subject with different poses. In the first image, the subject's head is inclined downwards, forcing their eyes to be upturned so they can look at the camera (eyes not shown to protect the identity of the subject). In the second image the subject is smiling, contorting their face making it difficult to assess the philtrum feature and the upper lip width and length. The third image presents a more neutral expression, with the subject's head directly facing the camera centralizing the subject's pupil. Inter-landmark distances were obtained using 23 of the landmarks used in the doll study (the tragions were excluded as they sometimes could not be viewed if the subject was not in the correct pose; the pogonion was also excluded, and a new landmark, the stomion was included). Two sets of 253 measurements per subject, obtained from two different image sets, resulted in a total of 2530 inter-landmark distances for 5 subjects. A comparison between the two data sets for each subject was made. Some error due to some degree of facial expression change between two sets of measurements was expected. Some inter-landmark distances, however, can be considered more stable positionally than others. For example, interpupillary distances would be expected to be more or less the same regardless of facial expression, while the mouth width would change dramatically if a subject was smiling. Reliability and normalized reliability were calculated, and the same grading scheme as described above was used.

Figure 4.6: A series of photographs of the same subject illustrating different poses.

4.4.2 Results

4.4.2.1 Inter-modality reliability of eye measurements

Table 4.5 shows the results of the comparison of eye measurements for 30 subjects. For ICD, 60% of the results were at least good, 24% moderate and 16% were poor in a comparison between stereo-photogrammetry and sliding caliper measurements. A comparison of IPD values revealed that 45.8% percent of the measurements were at least good, 33.3% were moderate and 20.8% were poor. For the PFL comparison, 56% of the values were at least good, with 32% moderate and 12% poor.

Table 4.5: Normalized inter-modality eye measurement comparison (Results are rounded and may add to slightly less or more than 100%)

	ICD (%)	IPD (%)	PFL (%)
Excellent	4.0	0.0	12.0
Very good	36.0	16.7	28.0
Good	20.0	29.2	16.0
Moderate	24.0	33.3	32.0
Poor	16.0	20.8	12.0

4.4.2.2 Effects of pose

The reliability of inter-landmark distances obtained from five live subjects using the stereo-photogrammetry imaging tool are shown in Tables 4.6 and 4.7. For subjects 1, 2, 3 and 5,; 79.8%, 86.6%, 70.4% and 83.4% respectively, of the inter-landmark distances were highly reliable. Only 41% of the inter-landmark distances measured for subject 4 were highly reliable. Subject 4 had the highest percentage of poor results at 25.3% but subjects 1, 2, 3 and 5, had poor reliability for 1.6%, 5.1%, 4.8% and 2.0% of their inter-landmark

distances, respectively. The normalized reliability for the 5 subjects followed the same trend.

Table 4.6: Pose assessment: reliability of stereo-photogrammetric measurements (Results are rounded and may add to slightly less or more than 100%)

	Subject 1 (%)	Subject 2 (%)	Subject 3 (%)	Subject 4 (%)	Subject 5 (%)
Highly reliable	79.8	86.6	70.4	41.1	83.4
Reliable	15.0	3.2	20.2	18.6	12.3
Moderately reliable	3.6	5.1	4.7	15.0	2.4
Unreliable	1.6	5.1	4.7	25.3	2.0

Table 4.8 shows some of the distances with poor precision on at least two subjects. The distance from the middle of the philtrum to the right crista philtri was poor for 4 of the 5 subjects. Three other distances, namely, pronasale to left subalare; pronasale to right subalare and philtrum centre to labiale superius, were poor in 3 subjects. It should be noted that a series of image sets was taken for each subject until one set with satisfactory pose was obtained. The above comparisons are therefore between the optimal image set and the next best set, and may overestimate the differences in inter-landmark distance that would be obtained if two optimal sets were compared.

Table 4.7: Pose assessment: normalized reliability of stereo-photogrammetric measurements (Results are rounded and may add to slightly less or more than 100%)

	Subject 1 (%)	Subject 2 (%)	Subject 3 (%)	Subject 4 (%)	Subject 5 (%)
Excellent	43.1	36.4	32.0	16.2	40.3
Very good	44.3	46.6	49.0	38.3	41.5
Good	8.7	7.1	11.9	25.7	9.1
Moderate	2.0	4.4	4.0	9.5	4.0
Poor	2.0	5.5	3.2	10.3	5.1

4.5 Two indirect anthropometry imaging tools

This section presents a comparison between craniofacial measurements obtained using the stereo-photogrammetry imaging tool and a commercial photogrammetry imaging system, the Vectra 3D (Wiltshire,UK). During the screening programme described in Section

Table 4.8: Pose assessment: poor inter-landmark distances

Distances	Subject 1 (%)	Subject 2 (%)	Subject 3 (%)	Subject 4 (%)	Subject 5 (%)
ls'R-alR		x			x
ls-sto			x	x	
Li-sto			x	x	
Phl-ls	x			x	x
Phl-ls'L	x				x
Phl-ls'R	x	x		x	x
Phl-sbalR		x		x	
psn-alL	x		x		
psn-enR		x	x		
psn-sbalL		x	x	x	
psn-sbalR		x		x	x
psn-sn		x		x	
sbalR-ls'L		x			x
se-n				x	x

4.4.1, both the stereo-photogrammetry imaging tool and the Vectra 3D were used to image 23 of the 254 subjects. To the author's knowledge, the reliability and precision of the Vectra 3D in craniofacial anthropometry has not been reported. Thus the term reliability was avoided in the comparison between the two systems and the term comparability was adopted instead. The Vectra 3D, however, has been used in a variety of facial anthropometry studies (Cox-Brinkman *et al.*, 2007; Hammond, 2007; Tobin *et al.*, 2008). As in the pose assessment study described in the previous section, some error was expected due to some degree of facial expression change between imaging sessions using the two systems.

4.5.1 Methods

4.5.1.1 Vectra 3D

The Vectra 3D is a commercial high-performance colour 3D facial surface imaging system. It is a two pod, six camera unit with a 180^o coverage that can acquire a full set of data in approximately 1.5 ms. The system projects a patterned light onto the scene and deformation is recorded by off-set cameras. Triangulation as described in Chapter 3 calculates a 3D mesh (MedEIM, 2004). Four of the cameras capture geometric information while two provide full colour texture images. Figure 4.7 shows an image of the Vectra 3D. Accuracy is claimed to be 100 μm root mean square (MedEIM, 2004). The

system can be configured to image subjects standing up or seated down. It comes with proprietary software to process and take measurements from the images. Zooming and rotation facilitate easier landmark location from the 3D composite image produced.

Figure 4.7: The Vectra 3D Scanner.

4.5.1.2 Study sample

Twenty-three subjects were imaged by both systems. Of the 23 image sets, 17 were of good enough quality to obtain measurements from both systems. All subjects were approximately twelve years of age (mean, 12.72 years; SD, 0.53) and twelve of the subjects had previously been diagnosed with FAS.

4.5.1.3 Comparability

Inter-landmark distances were obtained for 22 landmarks. The landmarks included were the same as those used in the doll study although the tragions were excluded. For each system one operator took two sets of measurements with at least 24 hrs between measurements. This resulted in four sets of 210 inter-landmark measurements (two sets obtained using the Vectra 3D and two sets obtained through stereo-photogrammetry) per subject. The comparability was assessed in the same way as reliability in Section 4.4.1. A

stereo-photogrammetry measurement less than 1 mm different from the Vectra 3D measurement was considered "very comparable"; between 1 and 1.5 mm, "comparable"; from 1.6 to 2 mm, "moderately comparable"; and anything more than 2 mm, "incomparable". The average (per measurement session) inter-landmark distances of the two systems were compared and the proportional error measure, REM, was calculated. The two sets of measurements obtained by the operator were averaged for the Vectra 3D measurements and for the stereo-photogrammetry measurements. These two sets of averaged values were then averaged to get a grandmean. The difference between the two sets was divided by the grandmean to give a measure of relative error, which is reported here as the normalized inter-modality comparability. Normalized inter-modality comparability scores less than 1% are deemed "excellent"; between 1 and 3.9%, "very good"; between 4 and 6.9%, "good", between 7 and 9,8%, "moderate" and finally anything above 10%, "poor", as described before. In addition, the systems were compared on a subject-by-subject basis providing a normalized inter-modality comparability index for each subject.

4.5.1.4 Assessing bias

It was assessed which system had a higher number of larger inter-landmark distances for each subject. The number of larger distances divided by the total number of distances as a percentage provided an index of bias.

4.5.2 Results

4.5.2.1 Comparability

A scatter plot of the averaged corresponding inter-landmark distances for the two systems is shown in Figure 4.8 (coefficient 0.9991, $p<<0$). There is a very high correlation in measurements between the two systems.

Table 4.9 shows the results of the inter-modality comparability between the two systems for all 17 subjects averaged over the two sessions. There is great variability in the per-subject comparability results especially in the percentage of incomparable results (incomparability range is from 15.2-53.3%). Three of the subjects' measurements show incomparabilty percentages greater than 50%. The results per subject show that there are two peak threshold categories, namely highly comparable and incomparable. Generally, however, the results reflect a poor comparibility between the two systems. Figure 4.9 shows a graph of mean inter-landmark differences (averaged over all subjects using

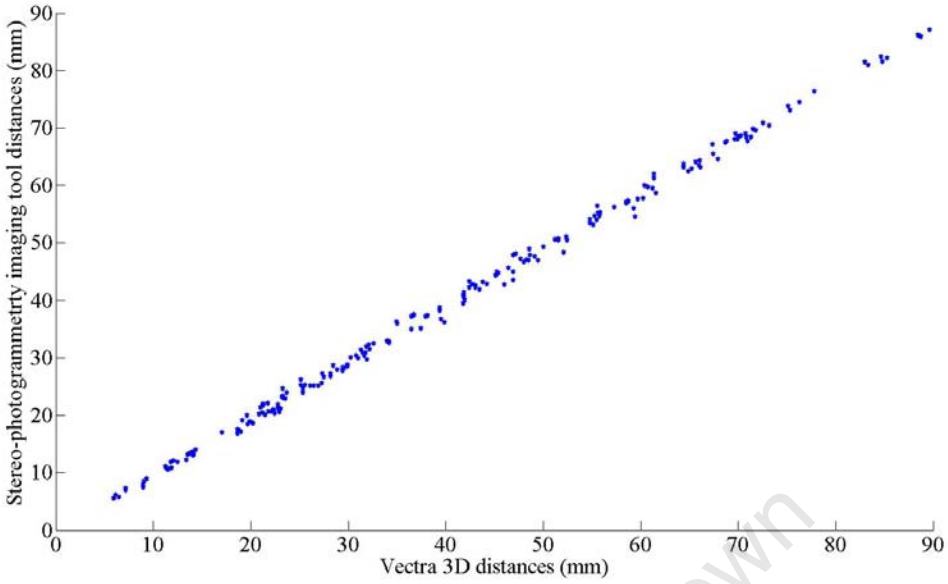

Figure 4.8: Correlation between the Vectra 3D and the stereo-photogrammetry imaging tool. The correlation coefficient was 0.9991, $p \approx 0$

the mean of two measurements for each system) vs. mean distances (averaged over all subjects and both systems). The mean differences range from 0 to 4.8 mm. This method of assessing differences between systems has been used by other researchers (Ghoddousi *et al.*, 2007; Kovacs *et al.*, 2006).

Table 4.10 shows the normalized inter-modality comparability for all 17 subjects. The results largely mirror the inter-modality comparability presented in Table 4.9. The range of poor comparability is 5.2-26.2%. The subjects (subjects 9, 12 and 16) with low comparability also show low normalized inter-landmark comparability. One threshold category, the very good category, has the highest percentage of measurements.

Table 4.9: Vectra 3D and the stereo-photogrammetry tool compared: inter-modality comparability (Results are rounded and may add to slightly less or more than 100%)

	Highly comparable (%)	Comparable (%)	Moderately comparable (%)	Incomparable (%)
Subject 1	41.4	19.1	17.1	22.4
Subject 2	31.0	14.3	18.6	36.2
Subject 3	39.5	16.2	14.8	29.5
Subject 4	39.5	12.9	16.2	31.4
Subject 5	36.7	13.3	13.8	36.2
Subject 6	37.6	11.0	14.8	36.7
Subject 7	61.4	10.5	10.0	18.1
Subject 8	58.6	18.6	7.6	15.2
Subject 9	26.3	10.5	10.0	53.3
Subject 10	30.5	11.4	10.5	47.6
Subject 11	47.1	20.0	13.3	19.5
Subject 12	26.7	8.1	12.9	52.4
Subject 13	41.0	14.3	16.7	28.1
Subject 14	38.1	20.5	14.3	27.1
Subject 15	33.8	13.8	10.0	42.4
Subject 16	23.3	14.8	11.0	51.0
Subject 17	28.6	17.6	12.9	41.0

The results show a strong bias as indicated in the last column of Table 4.10. The Vectra 3D measurements are consistently larger than the corresponding stereo-photogrammetry imaging tool distances. The combined average comparison shows an 83.8% bias towards the Vectra 3D having larger inter-landmark distances. This bias explains why the correlation between measurements was still good (Figures 4.8 and 4.9).

4.6 Chapter discussion and original contribution

The merits of a new technique can can only be assessed in light of the existing techniques (Majid *et al.*, 2005). The existing technique in craniofacial anthropometry is direct anthropometry and a comparison with this technique formed the basis for validating the stereo-photogrammetry imaging tool. Various protocols were followed in order to validate the measurement capabilities of the stereo-photogrammetry imaging tool. The first involved a comparison between direct anthropometry and the imaging tool by way of inter-landmark distances obtained from a marked doll. This method of assessing preci-

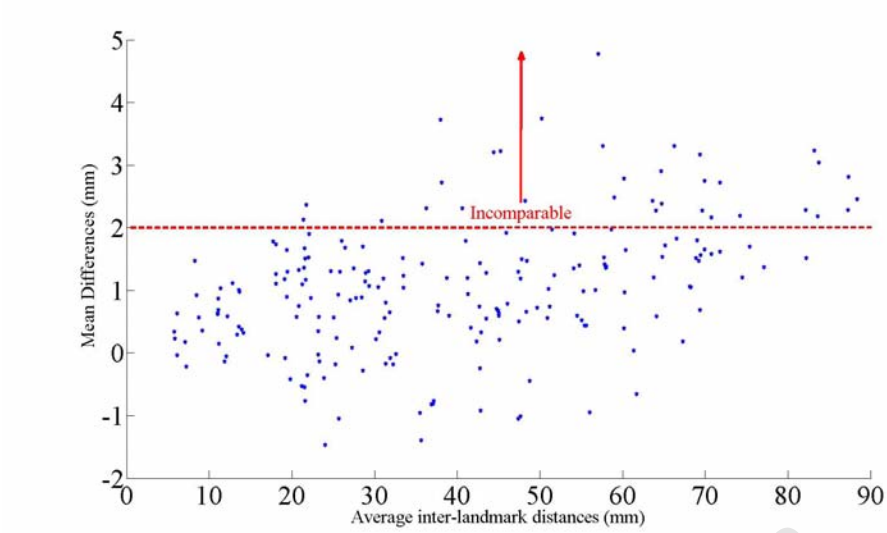

Figure 4.9: Mean differences between the systems vs. average inter-landmark distances. Each point represents a landmark. The differences for all the subjects were averaged to obtain the value for each landmark. The red line shows the 2mm threshold. Distances above this line are deemed incomparable.

sion and reliability has been widely reported in the literature (Kovacs *et al.*, 2006; Majid *et al.*, 2005; Weinberg *et al.*, 2006; Winder *et al.*, 2008) although precision and reliability have not been assessed with such a large number of inter-landmark distances. The use of a large number of distances allows for an assessment of reliability and precision with large coverage of the face and a greater range of distance magnitudes. Magnitude, especially, has previously been shown to affect error in measurement (Jamison & Ward, 1993).

The true values of the stereo- and direct-anthropometric measurements of the doll were unknown. Inter-modality reliability, rather than accuracy, was therefore used here because accuracy implies good knowledge of the quantity to be measured. Averaging over both operators and all measurement sessions for direct anthropometric measurements served to minimize error (Stallings & Gilmore, 1971) before adopting these measurements as reference values to be compared with the corresponding stereo-photogrammetric measurements. The direct anthropometric inter-landmark distances showed an acceptable inter-operator precision, which together with the REM precision, served to confirm the suitability of the direct anthropometric results as a basis for comparison with the stereo-photogrammetrically derived measurements. However, the difficulty of using manual instruments for small distances is evident here. Excluding the right alare to right

Table 4.10: Vectra 3D and the stereo-photogrammetry tool compared: normalized inter-modality comparability and bias (Results are rounded and may add to slightly less or more than 100%)

	excellent (%)	very good (%)	good (%)	moderate (%)	poor (%)	Vectra 3d > imaging tool [b] (%)
Subject 1	14.8	46.7	25.2	8.1	5.2	70.0
Subject 2	9.1	35.7	31.4	11.0	12.9	71.4
Subject 3	11.9	48.6	20.5	9.5	9.5	66.2
Subject 4	14.8	43.8	21.4	9.1	11.0	67.1
Subject 5	10.0	42.4	24.3	9.5	13.8	79.1
Subject 6	10.5	39.5	30.0	11.0	9.1	78.1
Subject 7	28.1	40.0	16.7	4.3	11.0	58.6
Subject 8	26.7	42.4	13.8	7.1	10.0	66.6
Subject 9	7.6	21.4	32.4	21.9	16.7	82.4
Subject 10	9.5	30.5	29.5	13.3	17.1	86.2
Subject 11	16.2	47.6	21.0	7.6	7.6	79.1
Subject 12	7.6	23.3	32.4	17.6	19.1	87.6
Subject 13	11.4	48.6	21.9	7.1	11.0	70.0
Subject 14	16.7	41.4	17.6	12.4	11.9	62.9
Subject 15	11.9	33.8	27.6	11.4	15.2	74.3
Subject 16	10.0	28.6	20.5	14.8	26.2	52.9
Subject 17	7.1	37.1	31.0	17.1	7.6	91.0

[b]Percentage of Vectra 3D measurements larger than the stereo-photogrammetric measurements

endocanthus distances, all the other poorly comparable inter-landmark distances are concentrated on the nose and upper lip area, and are the smallest distances measured here.

The inter-modality reliability results reveal that stereo-photogrammetric measurements are comparable to direct anthropometric measurements. All the measurements were at least "reliable" with 92.4% being "very reliable". The high degree of agreement between the two measurement techniques can largely be attributed to the use of a doll with labelled landmarks. The results are consistent with those in the literature (Kovacs *et al.*, 2006; Weinberg *et al.*, 2006; Winder *et al.*, 2008). There was an average bias of 0.4 mm towards larger stereo-photogrammetric distances. This would suggest a scaling error but there was no correlation between the magnitude of inter-landmark distances and that of measurement differences between the two modalities thus eliminating scale as a vector of error. The calipers used where not designed for anthropometry and this might have

caused some errors especially considering that some of the landmarks were on points of extreme curvature on a rigid doll making them difficult to measure directly.

The stereo-photogrammetry REM precision results were better than those of the direct anthropometric measurements. In addition, the direct anthropometric intra- and inter-operator precision results showed a higher degree of variability than the stereo-photogrammetry results. The literature supports the conclusion that 3D methods generally achieve greater precision than direct anthropometry (Meintjes *et al.*, 2002; Ras *et al.*, 1996a; Weinberg *et al.*, 2006). A possible explanation is that callipers and other direct methods are susceptible to parallax error as sometimes, in an effort to avoid subject discomfort, some measurements are made "in the air", i.e. with instruments hovering above the landmarks (Meintjes *et al.*, 2002). On the other hand, there are cases when instruments are directly in contact with the subject's skin, deforming it; this again may lead to errors. The use of a rigid labelled doll here, minimized both parallax and deformation type errors and served to diminish the spread of error allowing more focus on the inherent capability of the tool to make precise measurements. Inter-landmark distances about the upper lip and nose proved to be the most imprecise for stereo-photogrammetry. The reason might be that landmarks from these regions, specifically the left and right alare, subnasale and centre of the philtrum furrow, were difficult to locate in stereo-photogrammetry. Landmark identification has also previously been cited as a significant cause of error in anthropometry (Jamison & Ward, 1993). In addition the small magnitudes of the measurements are likely to be the cause of the larger errors.

Labelled landmarks do not completely eliminate error although magnification of images during measurements may minimize imprecision (Kovacs *et al.*, 2006). In the Australis software, it is possible to zoom in on labelled landmarks and this may have contributed to the highly precise and reliable results. The z coordinates were consistently less precise than the x and y coordinates. This in turn meant that the reliability results were affected as the distances were calculated using the same 3D coordinates used in precision assessment. The largest errors were along the z axis, as expected. Camera separation distance determines the relative resolution of that axis in relation to that of the other directions. Differing perspectives from the different cameras limited the degree of camera separation that could be achieved, affecting depth resolution. This aspect of photogrammetric design is always a trade-off between geometry and similarity of image perspective (to enable easier identification of features between scenes). Furthermore, an increase in camera separation would have resulted in increasing the size of the camera base which would have made the system both heavier and bulkier, impeding easy transportation to

remote sites. The median z-axis error values for both intra- and inter-operator precision (0.16 and 0.14 mm respectively) were deemed sufficiently low to be of no clinical relevance.

The eye measurement study showed mixed results. The purpose of the study was to compare the reliability of clinical measurements in a typical scenario with those derived using the imaging tool. Some error is expected when using physical instruments to take measurements of living subjects (Weinberg *et al.*, 2006). Measurements about the eyes were used because these were being taken as part of a FAS screening exercise using direct anthropometry, and additional discomfort to subjects could be avoided; however, eye measurements are usually the most error-prone especially in young subjects who may not be cooperative. In large-scale screening studies, the clinician's emphasis is in obtaining measurements for gauging percentiles within which measurements fall. Patients suspected of having facial anomalies are then referred for more detailed assessments including neurological and other evaluations such as the mother's history of alcohol consumption, before a final FAS diagnosis is made. The clinician's measurements are not generally repeated. This, coupled with the restlessness of young subjects, means that precise and reliable measurements are very difficult to achieve. This was one of the motivations for constructing this imaging tool: the intention was to assist in screening, reliably and precisely, a large number of subjects non-intrusively.

Accuracy of an earlier stereo-photogrammetric tool revealed no statistically significant differences between manual and stereo-photogrammetric measurements of PFL, while ICD and IPD differences were significant (Meintjes *et al.*, 2002); this assessment was done using a t-test, and the distribution of errors and their clinical acceptability was not presented. Thus, a direct comparison of our system with that used by Meintjes *et al.* (2002) is not possible. Our interpupillary distances produced the poorest comparison (<50% good agreement between the two modalities) followed by ICD (56% good agreement). Parallax error has been shown to play a large role in the errors produced in taking measurements of the eye that span the nasal bridge (Meintjes *et al.*, 2002); the difficulty in isolating the centre of the pupil and slight movements of the eyes have been suggested as the reasons for the disparity between the two modalities in IPD measurements. A source of error in the stereo-photogrammetry measurements was the difficulty in selection of corresponding points in the XY plane of each of the three 2D images (Douglas *et al.*, 2003b). Because of the potential depth difference, such error selection in 2D would result in over-or under-estimations in the resultant 3D calculations. This will always be a potential problem in this application of stereo-photogrammetry, since the operator is not landmarking on a single composite image. Eye measurements of subjects with slight epicanthal folds were

also difficult to obtain using stereo-photogrammetry. The zoom function in Australis, however, limited this source of error considerably.

The pose assessment reliability results in most instances indicated that stereo photogrammetry was highly reliable. The zoom function in Australis was once again critical in ensuring reliable landmark identification and good inter-landmark distance correspondence from different sets of images. One subject's results were much poorer than those of the rest. In the first set of images, the subject was smiling and their face was angled away from two of the cameras completely. In the second set the subject faced the central camera and had a more neutral expression. Most poorly corresponding distances were around the nose and upper-lip regions of the face (see Table 4.8) with subjects smiling or grimacing during image acquisition. As this is visually recognizable by the operator when reviewing the images, additional sets of images may be taken until an acceptable pose is achieved. Other subtle contortions of the face would be more difficult to perceive and therefore undetectable visually (for example, the sellion-nasion in Table 4.8).

The study comparing the Vectra 3D and the stereo-photogrammetry imaging tool, which was performed using data obtained in a screening exercise, produced mixed results. The correspondence assessment of averaged (per session) inter-landmark distances showed a high degree of correspondence as reflected by the high correlation coefficient (see Figure 4.8). This result was deceptive, however, as shown by the per subject comparability results (Table 4.9). The inter-modality differences ranged from 0 to 4.8 mm as shown in Figure 4.9, revealing large differences between the two systems in some measurements. The normalized inter-modality comparability results showed better comparability since the normalisation takes into account the size of the measurement. Stable landmarks, that is landmarks that do not change considerably with change of facial expression might have accounted for the majority of measurements considered comparable. Unstable landmarks, on the other hand, such as the corners of the mouth and eyes, whose positions vary considerably depending on facial expression, would then have accounted for the incomparable measurements. This effect was not investigated thoroughly partly due to time constraints and imaging conditions (the images were obtained during a screening programme and not in ideal laboratory conditions). Thus, since operators using the two different modalities only adhered to the image quality assurance protocol best suited to the modality they were using and no agreement on pose management was made before imaging, corresponding subjects often had different poses in the images taken using the different systems. All precision and reliability studies in the literature, using human subjects, have been performed under strictly controlled laboratory conditions, where the specific objective has

been the assessment of the comparability of measurements between the systems involved (Aldridge *et al.*, 2005; Chong & Mathieu, 2006; Ghoddousi *et al.*, 2007; Gwilliam *et al.*, 2006; Kovacs *et al.*, 2006; Majid *et al.*, 2005; Weinberg *et al.*, 2004, 2006).

The bias assessment showed a very high degree of bias. The Vectra 3D measurements were consistently greater than those derived from stereo-photogrammetry. In the literature, measurements obtained from 2D images are usually of smaller magnitude than those derived from 3D surface images (Ghoddousi *et al.*, 2007). In this case, where 3D methods are being compared, the degree of incomparability and consistently high bias suggest a scaling error might have played a greater role. Such bias, however, would not affect the use of the stereo-photogrammetric system as a screening tool, as bias would affect all subjects equally. A thorough comparison of the stereo-photogrammetry system, the 3D surface imaging system and direct anthropometry under laboratory conditions is required.

Chapter 5

Facial Shape Analysis

5.1 Introduction

Multivariate statistical procedures applied to distances, angles and distance ratios have been the traditional approach to morphometric analysis. These variables have the advantage of being invariant to position and orientation. Unless they are carefully selected, however, they may not be able to determine all the endpoints [1] of the features of interest and may exclude information about the shape structure of the object being analysed (Slice, 2005, 2007). Additional variables may need to be added to fix the relative positions of the endpoints, and this becomes impractical as the number of endpoints and variables increase. Landmark coordinates, however, encode all the information in any subset of distances or angles between them. The coordinate-based approach thus retains all geometric information from data collection through analysis and forms the basis of what are called geometric morphometrics. Geometric morphometric or shape analysis methods combine an explicitly geometric definition of shape with the flexible tools of multivariate statistics (Klingenberg & Monteiro, 2005), providing a basis for modelling shape differences both graphically and statistically (Hennessy *et al.*, 2002). The property of an object of greatest concern in geometric morphometrics is shape, which is inherently geometric and excludes colour and texture. Furthermore, shape is also invariant to location, scale and orientation (Slice, 2005), allowing objects to be compared on an equal platform regardless of scale or spatial orientation.

Shape analysis methods may play a valuable role after identification of facial landmarks that are considered important in the diagnosis of FAS. Facial shapes may be compared

[1] The endpoints are the points defining the distance measurement.

and averaged in terms of the relative positions of a set of landmarks (Clarren *et al.*, 1987; Streissguth *et al.*, 1991). In their report on the facial effects of fetal alcohol exposure, Clarren *et al.* (1987) utilized triangles defined by sets of three landmarks for analysing facial landmarks. The mean shapes of these triangles were compared between subjects more exposed or less exposed prenatally to alcohol. Recently, the author used images of seven-year old subjects obtained through stereo-photogrammetry in an attempt to characterize the shape variation of the FAS vs. the normal facial phenotype using geometric morphometrics (Mutsvangwa & Douglas, 2007). The stereo-photogrammetric tool consisted of a control frame with a pair of high resolution digital cameras (details in Meintjes *et al.* (2002)). A well-established approach was used in analysing shape variation using landmark data. Procrustes analysis followed by principal component analysis to reduce dimensionality and explore shape variability was shown to be a successful method for analysing the modes of facial variation between FAS and normal controls. Head circumference and midfacial features together with eye features, were found to play a role in differentiating between FAS and normal subjects. The study was limited, however, by a small study sample. In addition, the morphometric approach used in the study was not explicitly concerned with extrinsic factors affecting shape variability in biological specimens. Extrinsic factors are those that affect shape variation but are not the focus of the research question at hand. In the analysis of biological specimens using geometric morphometrics, size and age of the specimen are usually extrinsic factors in shape variability.

This chapter describes the use of geometric morphometrics, specifically facial shape analysis, in characterizing the facial phenotype associated with FAS. The chapter begins with an outline of the theoretical background of geometric morphometrics. Two approaches to facial shape analysis with respect to FAS are presented; one approach addresses the removal of extrinsic factors. The results from the two approaches and the limitations of both approaches are discussed.

5.2 Theoretical background

5.2.1 Definition of shape

Shape is defined as all the geometrical information that remains when location, scale and rotational effects are filtered out from an object (Dryden & Mardia, 1998). Two shapes can be compared by adjusting for size and superimposing one shape on the other using corresponding landmarks as references. The differences that remain are then due to

shape dissimilarity (Dryden & Mardia, 1998). Landmarks are divided into three categories (Dryden & Mardia, 1998):

- **Anatomical landmarks** are points assigned by an expert that correspond between organisms in some biologically meaningful way.
- **Mathematical landmarks** are points located on an object according to some mathematical or geometrical property, i.e. high curvature or an extreme point.
- **Pseudo-landmarks** are constructed points on an object either on the outline or between landmarks.

While orientation and location are easily removable nuisance parameters in geometric morphometrics, size may be of interest in biological morphometric studies. Size is an important component in the comparison of structures as it tends to dominate the variability between sexes, populations, species and even individuals (Slice, 2005). In morphometrics the term *form* is given to data containing only shape and size. Slice (2005) defines size as any positive, real-valued measure of an object that scales as a positive power of the geometric scale of the form. This definition of size, while very precise, may have ambiguous meaning in the presence of shape variation. An example is that the difference between two well defined points on an object constitutes a proper size measure, but for the same data set, different distances may indicate no change in size when objects differ in shape. Centroid size, geometric mean and linear distance are some of the size variables that are can be used in breaking down form into shape and size, depending on the research focus. Centroid size is the most commonly used size variable (Slice, 2005) and is defined as the square root of the sum of squared distances of landmarks in a configuration to their mean location (Slice, 2007).

5.2.1.1 Procrustes superposition

Procrustes superposition is a least-squares method that estimates the parameters for location and orientation that minimize the sum of squared distances between corresponding points on two configurations (Slice, 2005). A mathematical description of an n-point/landmark shape in k-dimensions is a concatenation of each dimension into a $(k \times n)$ vector. Establishing a common coordinate reference with respect to scale, position and rotation, aligns all the objects in a given set to obtain true shape representation. The Procrustes method for comparing shapes is a method used to fit all n landmark points

for N objects with optimal superimposing of landmarks. The method minimizes the distances between corresponding landmarks by removing all registration effects using some constraint. Various minimization constraints exist but the most popular is that which minimizes the sum of the squared distances between corresponding points (Halazonetis, 2004). For two shapes the following procedure performs the superposition of one shape to another:

1. The centroid size of each shape is calculated.
2. Each shape is normalized by dividing by the centroid size.
3. Each of the shapes is then aligned with respect to position at their centroids.
4. Each shape is then realigned with respect to rotational orientation about their centroids.

A basic illustration of a shape comparison is shown in Figure 5.1

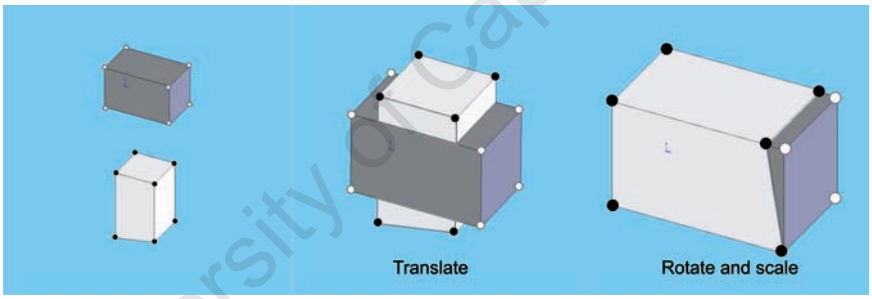

Figure 5.1: Shape comparison using Procrustes superposition.

In a multi-sample situation, the sample members are superimposed to a mean. However, there is usually no meaningful mean prior to the superposition. A solution is to use an iterative process to align shapes to a changing mean shape. Any of the shapes in the sample can be used as an initial mean and the others aligned to it. A new mean from the aligned configurations may then be calculated as the arithmetic average location of individual landmarks in the sample and scaled to unit centroid size. This process is guaranteed to produce monotonically decreasing sum-of-squared deviations of the sample configurations around the estimated mean (Slice, 2005). Convergence may be declared when the mean shape does not change significantly within an iteration (Stegmann &

Gomez, 2002). This process is called Generalized Procrustes analysis (GPA). Figure 5.2 illustrates GPA for three facial landmarks from a sample of 2D faces.

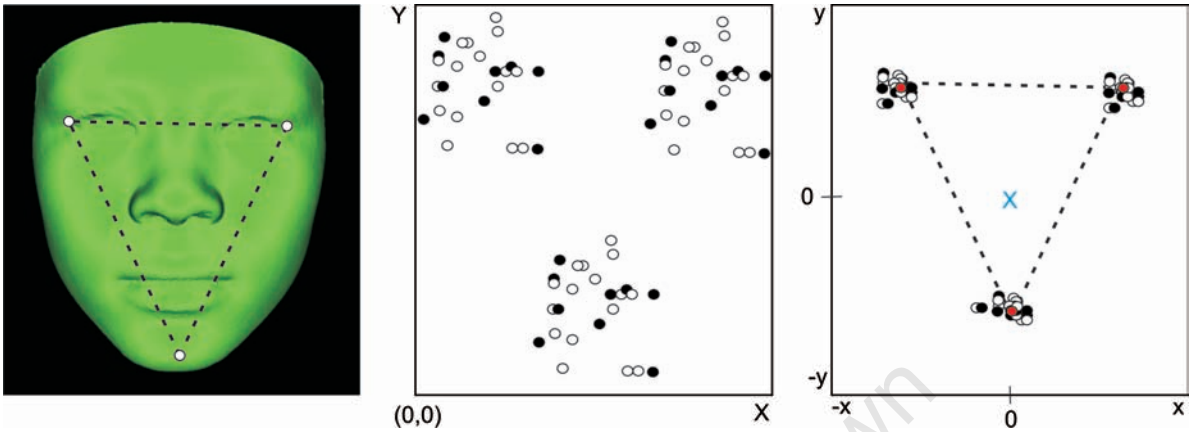

Figure 5.2: Generalized Procrustes analysis (GPA). The coordinates of corresponding landmarks may be taken from several samples of faces as shown for one sample in the first figure. The dashed line defines the shapes being compared, which are made up of three landmarks, the two exocanthi and the gnathion. The second figure shows a collection of such samples before superposition. The third figure shows the sample after superposition. The red circles are the coordinates of the mean shape after the GPA. The blue cross is the centroid of the mean shape and is positioned on the origin after GPA.

Mathematically, for a sample of landmark configurations S_i, where $i = 1, \ldots N$ and N is the number of specimens, $\overline{\alpha}$, the Procrustes mean shape after convergence can be estimated by:

$$\overline{\alpha} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \tag{5.1}$$

and gives the Procrustes mean coordinates ($\overline{\alpha}_{jx}$,$\overline{\alpha}_{jy}$,$\overline{\alpha}_{jz}$) where $j = 1, \ldots n$ (n is the number of landmarks). The full Procrustes fit coordinates S_i^P are found by fitting each shape S_i to the Procrustes mean $\overline{\alpha}$ using, for example, a least squares method (Halazonetis, 2004) and thus each S_i^P has coordinates: (S_{ijx}^P,S_{ijy}^P,S_{ijz}^P). Procrustes residuals S_i^{PR} (Robinson *et al.*, 2001) are the difference between the full Procrustes fit coordinates and the Procrustes mean coordinates, and are here represented as: (S_{ijx}^{PR},S_{ijy}^{PR},S_{ijz}^{PR}). Procrustes residuals have statistically useful properties (McIntyre & Mossey, 2003) and can

be used in principal component analysis to explore shape variability (Hennessy *et al.*, 2002).

5.2.1.2 Shape spaces and multivariate analysis

The Procrustes distance of a shape configuration is the square root of the sum of squared differences of corresponding landmark coordinates between the Procrustes aligned object and the computed configuration mean after convergence. This is a non-Euclidean distance and a set of all such distances for corresponding configurations in a sample reside on the surface of a hyper-hemisphere. Since this space is non-Euclidean, however, standard parametric multivariate statistical methods are not directly applicable to the data in it. An important feature of this space is that most biological shape variability of interest to researchers is concentrated in a small area (Slice, 2005), and when variation in shape is reasonably small, a projection of the Procrustes fit coordinates into linear tangent space can be made to produce Procrustes tangent coordinates. Tangent space projection makes linear assessment of the shape space possible and thereafter standard multivariate methods may by applied. The Procrustes residuals can be used as a good approximation to the Procrustes tangent coordinates.

5.2.1.3 Principal component analysis

Procrustes residuals may be used to explore shape variability (Hennessy *et al.*, 2002). The variability of shape around the mean can be viewed as scatter plots or connected scatter plots for easier visual inspection. The position of each point can vary along all orthogonal axes. Mathematically, $(k \times n)$ Procrustes residual variables can describe the variability of the object. Various statistical procedures can be carried out to investigate shape variation. One of them is principal component analysis (PCA). PCA is a procedure that can decrease or summarize the $(k \times n)$ variables into uncorrelated linear combinations of these variables. The r^{th} principle component $PC(r)$ is given by:

$$PC(r) = \sum_{j}^{n} (\gamma_{rjx} S_{jx}^{PR} + \gamma_{rjy} S_{jy}^{PR} + \gamma_{rjz} S_{jz}^{PR}) \tag{5.2}$$

where the weights γ_{rjx},γ_{rjy} and γ_{rjz} describe directions of variation in each Procrustes coordinate about the mean shape (Robinson *et al.*, 2001). A set of the same shape class of biological objects will always have some degree of inter-point correlation (Stegmann and Gomez 2002), especially since they belong to the same biological entity (Halazonetis,

2004). There therefore exists a shape representation between points that accounts for the correlation between points. This can be used to reduce dimensionality in principal component analysis. There are as many principal components as there are landmark coordinates. The principal components have the following properties (Halazonetis, 2004):

- All components are orthogonal to each other and so are statistically unrelated.
- Part of the variability of the sample is represented by each component in decreasing order, starting from the largest variability being represented by the first component and the second largest variability by the second and so on.
- Every component is a linear combination of the original variables.

One of the reasons to perform shape analysis is to ascertain whether there are any shape differences within a population of shapes. Some of the shape variation represented by the principal components may account for shape differences between clusters of shapes within the sample being investigated. To determine which principal components have greater discrimination between such clusters, if they exist, a discriminant analysis of principal component scores can be done. Discriminant function analysis or discriminant analysis is used to determine which variables (usually continuous) discriminate between two or more naturally occurring groups. The components found to have the greatest discriminating power can be further analysed using scatter plots of principal scores and shape warping. Scatter plots provide a useful visual illustration of where each shape lies along the direction of a chosen principal component. To visualize the pattern of shape variability represented by each principal component, the mean shape can be warped by moving points according to the weights on the principal component. The decreasing order of significance of the components in accounting for variability means that only some need be retained to account for a significant part of shape variability. The mean shape configuration of a sample has, by definition, all principal components equal to 0. For $PC(r)$, shapes with the following coordinates can be plotted: $(\overline{\alpha}_{jx} + c\sqrt{\lambda_r \gamma_{rjx}}\ ,\ \overline{\alpha}_{jy} + c\sqrt{\lambda_r \gamma_{rjy}}\ ,\ \overline{\alpha}_{jz} + c\sqrt{\lambda_r \gamma_{rjz}})$. Here λ_r is the variance captured by $PC(r)$. The value c can range about the mean configuration by some standard deviation in the negative or positive direction of $PC(r)$. Shape is considered to be a continuous variable of smoothly varying patterns and not a discrete variable.

5.2.1.4 Symmetry

Bilateral symmetry exhibited by some biological organisms can result in ill-conditioned covariance matrices resulting from linear dependences among landmarks due to symmetry. Two types of bilateral symmetry exist in biological organisms (Klingenberg *et al.*, 2002). Matching symmetry is the description given to the symmetry where two separate copies of a structure, for example a wing in insects, are present on both sides of the body. The other type of symmetry is objective symmetry. Here a structure is symmetric in itself because the mid-sagittal plane or median axis passes through it, as in the face (see Figure 5.3).

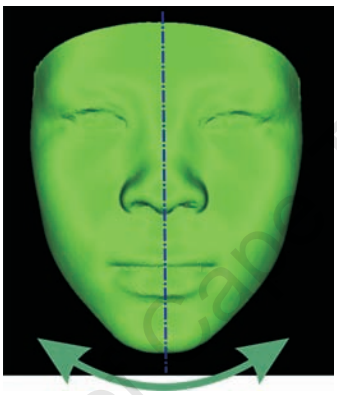

Figure 5.3: Objective symmetry

Most biological organisms do not exhibit complete symmetry but show some asymmetrical variation in the position of some features. Landmark configurations with objective symmetry can be analysed by first partitioning total shape variation into components of symmetry variation and asymmetry. Figure 5.4 illustrates the above in that the distorted face below exhibits asymmetry and symmetry simultaneously. The face is asymmetrical about the midline but all points on the median axis are symmetric on themselves. Klingenberg *et al.* (2002) proposes a method to partition total shape variation into symmetric variation and asymmetry. Each shape configuration can be reflected along its median axis. The landmark labels of the reflected copy are then reassigned to correspond to those of the original copy. Both the reflected copy and the original copy are included in a GPA to obtain a perfectly symmetric Procrustes average of the original configuration.

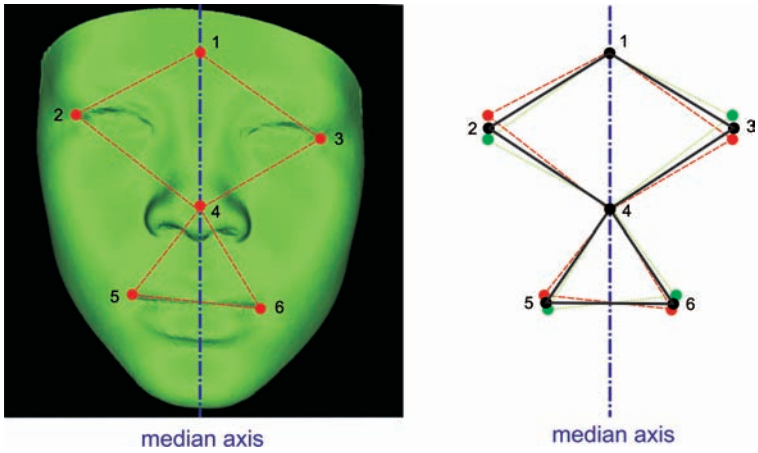

Figure 5.4: Symmetry correction. The left image shows an asymmetric face. The right image shows the original shape configuration (red dashed line) reflected along its median axis. The landmark labels of the reflected copy (green dashed line) are then reassigned to correspond to those of the original copy. Both the reflected copy and the original copy are included in a GPA to obtain a perfectly symmetric Procrustes average of the original configuration (black solid line).

5.3 Facial shape analysis: introduction to two studies

This section introduces two facial shape analysis approaches to characterize the facial phenotype associated with FAS. One approach uses GPA to align facial landmark configurations and PCA to explore facial shape variation as previously reported in Mutsvangwa & Douglas (2007). In addition, pattern recognition methods, specifically feature selection and classification methods are employed on the principal components to determine which are most discriminatory between normal and FAS-affected subjects. The characteristic facial phenotype for FAS can then be deduced from a visual presentation of the shape variation described by the most discriminating principal components. The other approach attempts to correct for the effects of size and age using regression after GPA. Discriminant function analysis is then used to assess the age- and size-invariant facial shape variation associated with FAS and classify subjects using cross-validation.

5.3.1 Study sample

Thirty-four subjects, in two age groups, five and twelve years old, were included in the study (see Table 5.1). Subjects were of mixed ancestry (people from intermarriage of black African populations, European-origin whites, and some Asians) (May *et al.*, 2007).

The study sample was chosen from a larger population of children imaged as part of a large-scale FASD screening program in the Northern Cape Province of South Africa. The screening process was through sequential visits to suspected high prevalence areas under a Foundation of Alcohol Related Research (FARR) initiative. FARR is a non-governmental organisation in South Africa conducting research in alcohol related medical problems in South Africa.

Children were screened for signs of growth retardation based on height, weight and head circumference and their facial characteristics were evaluated for FAS by a dysmorphologist using the Hoyme Diagnostic Guidelines (Hoyme *et al.*, 2005). A child with a diagnosis of FAS had sufficient dysmorphology, was approximately two standard deviations below the mean on either verbal or non-verbal intelligence quotient tests, had substantial behavioral problems as measured by the Personal Behavior Check-list (PBCL-36)[1] and had confirmation of prenatal alcohol exposure.

The study sample of five-year olds represents an age-matched subset obtained during visits to high risk areas. The term "'normal controls"' is used to refer to subjects not flagged for growth retardation and not presenting any of the FAS facial characteristics as determined by a dysmorphologist. Because the small samples are not the same as comparing true population norms these subjects should strictly be called controls. The normal subjects were chosen in order to match them in age and size to the FAS group.

The study sample used here had a greater number of older subjects with FAS and conversely a greater number of younger control subjects. The mean age for the control subjects was 8.0 $\pm$ 3.6 years. The mean age for FAS subjects was 10.9 $\pm$ 3.6 years. The disparity in means was significant ($T^2 = 2.44$, p=0.01). For the two age samples separately, the differences in mean and standard deviation between FAS and control subjects were not significant (see Table 5.1).

Twenty-six landmarks were obtained for each of the subjects (see Figure 5.5). These included landmarks describing features traditionally assessed in FAS diagnosis, namely those defining the palpebral fissures, the upper vermillion boarder and the philtrum (Astley & Clarren, 1995, 1996, 2000; Clarren *et al.*, 1987; Moore *et al.*, 2001, 2002, 2007; Mutsvangwa & Douglas, 2007). These landmarks are also generally included in the description of the three most severe of the four fetal alcohol spectrum disorders (FASD)

[1]This is a scale that measures the behavioral characteristics of FAS, regardless of age, race, sex, or I.Q., consisting of 36 items pertaining to several areas of functioning: academic performance, social skills and interactions, bodily functions, communication and speech, personal manner, emotions, and motor skills (May *et al.*, 2007).

Table 5.1: Descriptive statistics

		Five-year olds	Twelve-year olds	All subjects
Normal controls	Mean (Std)	5.4 (0.5)	12.7 (0.5)	8.0 (3.6)
	Number of subjects	11	6	17
FAS	Mean (Std)	5.1 (0.4)	12.7 (0.5)	10.9 (3.4)
	Number of subjects	4	13	17
Difference in mean age	T^2	-1.17	-0.01	2.44
	p	0.14	0.50	0.01

(Manning & Hoyme, 2007). The definition of FASD is described in Section 2.1. Landmarks defining midfacial hypoplasia were also included on the midline of the face. This facial anomaly can be assessed using a subset of linear distances only (Moore *et al.*, 2001, 2002, 2007) but has been assessed spatially in 3D (Mutsvangwa & Douglas, 2007). The reliability of the imaging tool and the landmarking of images have previously been assessed using a doll and human subjects, as described in Section 4.3.

5.3.2 Shape data

Each subject's set of 3D landmark coordinates constitutes a form (shape and size) configuration. In the analysis, symmetry correction and GPA were applied to the 3D coordinates, and the Procrustes residuals formed the shape data. These data were analysed using two different approaches as described below. The two age categories allow an explorative investigation into the differences in the FAS facial shape at different ages and provide an assessment, qualitatively and quantitatively, about the age at which the deviation of the FAS facial shape from the norm is more pronounced. The assessment of the FAS facial shape at five was motivated by the consensus in the research community that FAS is easier to diagnose below the age of 10 years (Astley & Clarren, 1995). An investigation into facial shape difference for twelve-year olds was motivated by its potential to demonstrate if the facial anomalies associated with FAS are detectable after the onset of puberty. The morphometrics software MorphoJ version 1.00h (Klingenberg, 2008), and a Matlab toolbox for pattern recognition called PRtools version 3 (Duin, 2000) were used in the analyses.

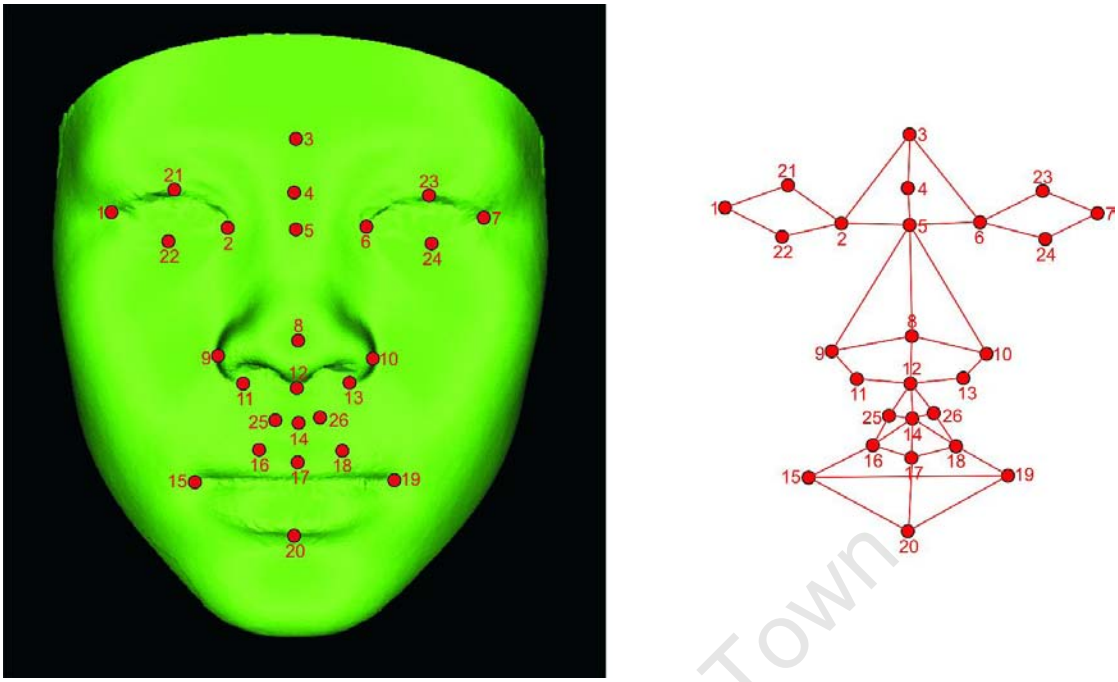

Figure 5.5: The landmarks included in the analysis. 1. exR -right exocanthion, 2. enR-right endocanthion, 3. g-glabella, 4. n-nasion, 5. se-sellion, 6. enL-left endocanthion, 7. exL-left exocanthion, 8. psn-pronasale, 9. alR-right alare, 10. alL-left alare, 11. sbalR-right subalare, 12. s-subnasale, 13. sbalL-left subalare, 14. Phl-centre of philtrum furrow, 15. chR-right cheilion, 16. ls'R-right crista philtri, 17. ls-labiale superius, 18. ls'L-left crista philtri, 19. chL-left-cheilion, 20. li-labiale inferius, 21. umeR-right upper mid eye ridge, 22. umeR-right lower mid eye ridge, 23. umeL-left upper mid eye ridge, 24. umeL-left lower mid eye ridge, 25. phmR-mid right philtrum ridge and 26. PhmL-mid left philtrum ridge. The wire frame shown on the right figure is used for visualization purposes only and does not affect the analysis.

Every analysis presented below involved an initial GPA. To assess if there are outliers in the data after GPA, the MorphoJ software provides a diagram with the cumulative distribution of the distances of individual specimens from the average shape of the entire sample. A deviation from the curve expected for a multivariate normal distribution fitted to the data shows the extent to which an individual is an outlier, based on Procrustes or squared Mahalanobis distance, depending on the relationship between the dimensionality of the data and the number of specimens in the dataset. This provides a visual inspection of outliers in the data.

5.4 Facial shape analysis using principal components

5.4.1 Methods

5.4.1.1 Discriminant function analysis of Procrustes residuals

To assess the difference in mean shapes after GPA, a discriminant function analysis of the Procrustes residuals was done for each of the age categories and for all ages combined. Discriminant function analysis aims to determine whether groups differ with regard to the mean of a variable, and then to use that variable to predict membership to a group (Hill & Lewicki, 2007). In this case, the variables were the Procrustes residuals and group membership was the diagnosis of FAS or normal. This was done for each of the age categories and for all the subjects combined.

5.4.1.2 Principal component analysis

In this study, principal component analysis was performed on the covariance matrix of symmetric Procrustes residuals after GPA. Three separate analyses were done, one for each of the age group categories and one with all the subjects combined. After the principal component analysis, the principal component scores along each principal component for each subject were computed. Principal component scores are computed as the vectors of deviations of the observations from the sample mean, multiplied by the vectors of principal component coefficients or eigenvectors (Klingenberg, 2008). The number of principal components to retain is determined by the research question at hand. The Jolliffe cut-off value for eigenvalues produced in a principal component analysis, is a method to show informally how many principal components should be considered significant (Jolliffe, 2002). This method retains the principal components associated with the covariance matrix, that have eigenvalues greater in magnitude than the average of all the eigenvalues. Other criteria exist for determining this cut-off; Cangelosi & Goriely (2007) provide a good summary of the methods. In the present study all principal components were retained for further analysis despite their apparent insignificance. Dimensionality reduction, the usual reason for a principal component analysis, was performed later through feature selection as described below. The motivation for retaining all the principal components was that the FAS facial phenotype may present very subtly and thus some principal components which might not be significant in overall shape variation might still be quite discriminatory between the FAS and normal shapes. Table A.1 in Appendix A shows the principal components from each of the three analyses.

5.4.1.3 Feature selection

Feature selection is used to select a subset of relevant variables, thereby reducing the dimensionality of the data space, such that maximum discrimination of groups or targets can be achieved according to some criteria. The usual criterion is the minimization of classification error and computational effort for classifier algorithms. Mathematically, feature selection translates to choosing a subset of k features out of n that guarantees the lowest classification error, which is an optimization problem, barring an exhaustive search for the subset of features (Frank, 2002). In this study, the principal component scores along each principal component were the feature variables. The targets were coded, 1 for FAS affected subjects and 0 for normal controls. Branch-and-bound feature selection was used to obtain the subset of principal components that best discriminated between the two groups, FAS and normal controls. The branch-and-bound algorithm is a top-down procedure, beginning with a set of p variables, and constructing a tree by deleting variables successively (Webb, 2002). It relies on an important criterion for feature selection, which is that, for two subsets of variables, X and Y:

$$X \subset Y \Rightarrow J(X) < J(Y) \tag{5.3}$$

where J is a measure of separation between two sets of data. This relation says that evaluating the feature selection criterion on a subset of variables of a given set yields a smaller value of the feature selection criterion, a property referred to as the *monotonicity property*. At each point of searching the tree, a bound is computed of the best solution possible[1] in the current sub-tree. Promising nodes of the tree are retained and expanded, whereas, due to the monotonicity property, nodes for which the lower bound is lower than the best solution so far are pruned (Frank, 2002). In the software package PRtools, the separation measure maximized was the sum of the squared Mahalanobis distances between targets in the data space. The output from the feature selection was the optimum subset of principal components ranked in order of decreasing discriminatory ability. Principal component scores from the first two[2] principal components in the ranking were used in classification. The orthogonality property of principal components guarantees independence, allowing individual principal components to be chosen based on their individual classification errors, thus a subset of features can be pooled based on their individual classification error performance.

[1]The best solution is the highest value of the measure of separation between classes.

[2]For the three analyses, the optimal subset was consistently two principal components.

5.4.1.4 Classification

Four classifiers that were used to classify the subjects are briefly described below (Duin, 2000):

- **Nearest mean classifier**. This classifier computes the mean of each group and allocates group membership based on proximity to group mean.
- **Fisher classifier**. This classifier finds the linear discriminant function between the groups in the dataset by minimizing the errors in the least square sense.
- **Quadratic Bayes classifier**. This classifier works by finding the covariance and mean of a group and using that to create a decision boundary.
- **Log linear classifier**. This is a linear classifier that maximizes the likelihood criterion using the logistic (sigmoid) function.

Due to the small sample size, leave-one-out cross-validation classification was used. In each trial all but one subject were used in the training of a classifier. The subject left out was then used in testing a classifier resulting in 33 trials for each classifier. A visualization of the shape variation described by the most discriminatory principal components provided a qualitative description of facial shape changes between FAS and normal controls.

5.4.2 Results

5.4.2.1 Discriminant function analysis

Table 5.2 shows the results of the discriminant function analysis of Procrustes residuals. There is no statistically significant difference in mean shapes between FAS and normal controls for each of the age group categories and for all the subjects combined. The shape differences were thus too subtle to compare mean shapes after GPA alone.

Table 5.2: Discriminant function analysis of Procrustes residuals for FAS and normal control subjects.

	Five-year olds	Twelve-year olds	All subjects
T^2	18.20	20.65	309.78
p-value (parametric)	0.99	1.00	0.92

Figures 5.6, 5.7 and 5.8 below show the shape changes associated with the principal components chosen from the feature selection for each of the analyses. The change is from the mean shape in the positive and negative directions of the principal component by a scale factor. The scale factor is the magnitude of the shape change as a Procrustes distance. A scale factor of 0.1 corresponds to a change of the principal component score of 0.1 units in the positive direction while a scale factor of -0.1 corresponds to -0.1 units in the negative direction. All variation in Figures 5.6, 5.7 and 5.8 is described starting from a negative scale factor, past the mean shape towards a positive scale factor. The scale factors were determined by the variation about the mean of subjects in the scatter plots of the principal components (see Figures A.1, A.2 and A.3 in Appendix A for scatter plots of the principal components of the five-year old, twelve-year old and combined group, respectively).

Figure 5.6 shows the shape variation described by principal components 2 and 5 (selected by feature selection) for the five-year old sample. The scale factor is 0.05 in either direction of the mean. Principal components 2 and 5 account for 15.8% and 6.9% of shape variability, respectively (Table A.1 in Appendix A). The shape variation in principal component 2 shows a large reduction in the size of the eye orbits. The glabella also changes position dramatically from a more inferior to a more superior position. There is a also a thinning of the upper lip, increase in philtrum length and flattening of the midface. The shape variation in principal component 5 includes a medio-lateral moevement of the inner canthi, thinning of the both the upper and lower lip and flattening of the midface. The FAS subjects in the scatter plot of principal components 2 and 5 lie in the upper right hand corner (Figure A.1 in Appendix A). It can be concluded that the right hand shapes for principal components 2 and 5 belong to the FAS group and those to the left are the normal control faces.

Figure 5.7 shows the shape variation described by principal components 6 and 15 for the twelve-year olds (scale factor is $\pm$ 0.03). Principal component 6 and 15 account for 4.9% and 0.7% of shape variability, respectively. Shape variation, in the sagittal view, of principal component 6 shows a ventro-dorsal change in the relative position of the nasion and seillion with glabella changing in the opposite direction. There is also evidence of the nasal protrusion diminishing. The shape variation described by principal component 15 is limited to a slight change in the relative position of the nasion mirroring that landmark's change in principal component 6. Generally, the changes in shape are very small. The FAS subjects in the scatter plot of principal components 6 and 15 lie in the upper left hand corner (Figure A.2 in Appendix A). It can be concluded that the left hand shape

for principal component 6 belongs to the FAS group and that to the right is the normal control face. For principal component 15 the shape to the right is the FAS face and that to the left is the normal control face.

Figure 5.8 shows the shape variation described by principal components 1 (scale factor is ± 0.05) and 15 (scale factor is ± 0.01) for both ages combined. Principal components 1 and 15 account for 24.6% and 1% of shape variability, respectively. As mentioned above, the dominant extrinsic factor in morphometrics is usually size. In addition, without size correction, the first principal component after a GPA is usually responsible for size differences (Swiderski, 2003). To test for this, centroid size was regressed on principal component 1. Size correlated poorly with principal component 1 and only predicted 12.1% of variation (R= 0.35 p=0.04). Shape variation, in the coronal view, for principal component 1 shows a significant lengthening of the palpebral fissures and up-slanting of the eyes. In the sagittal view there is a large dorso-ventral and infero-superior movement of the glabella and nasion and sellion, although the nasion shows a greater change in relative position. A significant difference can be observed in the vertical length of the midface, from the seillion to the labiale superius. Significant thinning of both the upper and the lower lip is also evident. Shape variation described by principal component 15 includes slight changes in the relative position of the nasion and pronasale and slight changes in the mouth width. The FAS subjects in the scatter plot of principal components 1 and 15 lie in the lower left hand corner (Figure A.3 in Appendix A). It can be assumed that the left hand shapes for principal components 1 and 15 belong to the FAS group and those to the right are the normal control faces.

The classification accuracy of the four pattern recognition algorithms for the three analyses is shown in Table 5.3. The focus of the analyses was identifying FAS subjects so the definition of accuracy as it is used here is:

$$accuracy = \frac{no\ of\ true\ FAS\ positives\ +\ no\ of\ true\ control\ negatives}{total\ number\ of\ subjects} \times 100 \quad (5.4)$$

Figure 5.6: The shape variation described by principal components 2 (top) and 5 (bottom) for five-year olds. The left figures show shape change in the negative direction from the mean (black solid line). The change in shape is shown by the blue dashed line. The figures to the right show shape change in the positive direction. The shape change is shown by the dashed red line. The numbers are the scale factors.

Figure 5.7: The shape variation described by principal components 6 and 15 for twelve-year olds. The left figures show shape change in the negative direction from the mean (black solid line). The change in shape is shown by the blue dashed line. The figures to the right show shape change in the positive direction. The shape change is shown by the dashed red line. The numbers are the scale factors.

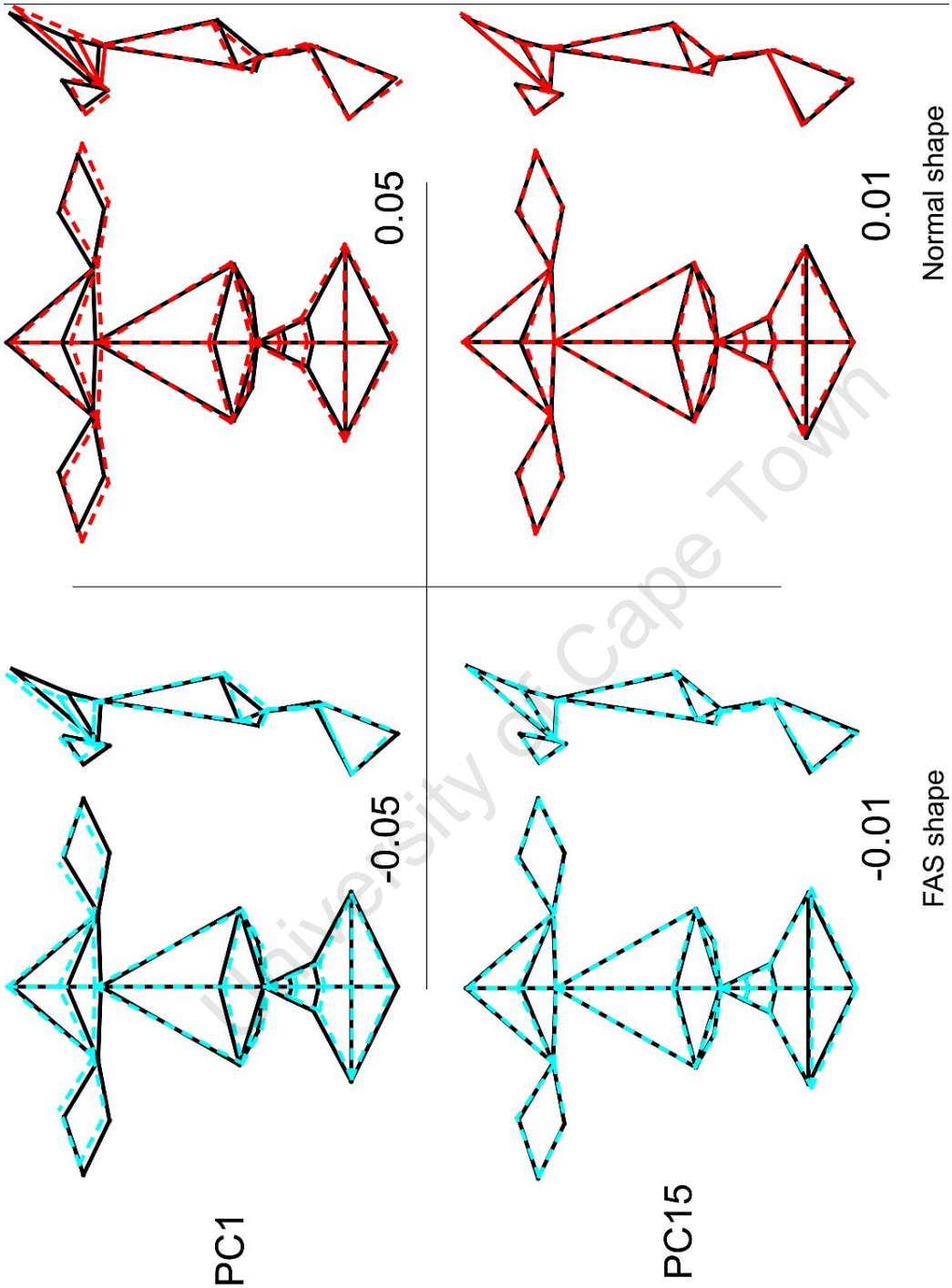

Figure 5.8: The shape variation described by principal components 1 and 15 for all subjects. The left figures show shape change in the negative direction from the mean (black solid line). The change in shape is shown by the blue dashed line. The figures to the right show shape change in the positive direction. The shape change is shown by the dashed red line. The numbers are the scale factors.

Table 5.3: Accuracy of classification for all analyses.

Classifier	[a] Five-year olds PC 2 and 5 %	Twelve-year olds PC 6 and 15 %	All subjects PC 1 and 15 %
Nearest Mean	80	74	76
Fisher	87	74	74
Quadratic Bayes	87	79	62
Log Linear	87	74	76
Average	85	75	72

[a]PC indicates principal components obtained in the feature selection.

5.5 Facial shape analysis with regression for age and size correction

5.5.1 Methods

In studies concerned with shape only, extrinsic factors affecting shape variability can be removed. Extrinsic factors are those that affect shape variation but are not the focus of the research question at hand. In the analysis of biological specimens, using geometric morphometrics, size is usually an extrinsic factor in shape variability. Shape variability may also depend on the age differences in the specimens being compared. Although age and size are expected to be related, some shape change may occur as individuals age. Regression can be used to correct shape data for extrinsic factors contributing to shape variability. The following description of how regression may be used in geometric morphometrics is taken from Klingenberg (2008).

Regression separates the component of variation in the dependent variables that is predicted by the independent variables from the residual component of variation, which is uncorrelated with the independent variable. Regression can be used, for instance, in correcting for size which may be an extrinsic factor in shape variability. This decomposition is shown in Figure 5.9a for a single data point (black dot). The deviation from the sample average (hollow dot) in the direction of the shape variable (vertical; the dependent variable) is divided into a predicted and a residual component. The predicted component can be computed from the slope of the regression line (bold oblique line) and the deviation of the data point from the mean in the direction of centroid size (horizontal; the independent variable). The residual is the difference of the total shape deviation of

the data point from the mean and the predicted component. By removing the predicted component and focusing on the residuals, what is left is data that is uncorrelated with the independent variable. In the case of a two-group analysis, if the regression slopes in the groups are the same, a pooled regression can be performed. This method uses the regression slopes within samples to separate the predicted and residual components of variation in the dependent variables as shown in Figure 5.9b.

In this study, contributions to shape variation from age and size were corrected for after GPA. To correct for age, the shape data were regressed onto subject age for all subjects using pooled within-group regression. This separated the components of shape variation that were predicted by age from those that were independent of age. A permutation test[3] for the null hypothesis of independence of shape from age was performed. The residuals from the regression were therefore assumed to be independent of age and constituted the new shape variables for further analysis.

These new shape data were then also regressed on centroid size using pooled-within group regression. The overall residuals from the tandem regression of shape on age and age-corrected shape on centroid size were retained for further analysis. Discriminant function analysis of the size- and age-invariant shape data was then used to analyse the shape differences between the control subjects and the FAS subjects for 1) the five-year old subjects and 2) the twelve-year old subjects and 3) all subjects combined. Because we are dealing with shape variables, the average shape differences may be assessed visually. In addition, the performance of the discriminant function in predicting group membership may be assessed using leave-one-out cross-validation. The FAS face for younger individuals is expected to be different from that of older individuals. To assess this difference, the five-year old FAS sample was compared to the twelve-year old FAS sample. The same was done for the control subjects at the different ages. Although the normal face is expected to change with growth, it was hoped that a comparison of the differences within the FAS and control groups at different ages would reveal any differences in the development of

[3]A permutation test involves the shuffling of observed data in order to assess the uniqueness of an observed outcome. A suitable test statistic is computed on the observed data associated with a collection of objects. The data are permuted over all possible combinations of the objects and the test statistic is computed for each combination. The null hypothesis specified by the randomization implies that each combination is equally likely to occur. The p value is the proportion of combinations with test statistic values as extreme or more extreme than the test statistics computed on the original combination (Mielke & Berry, 2007).

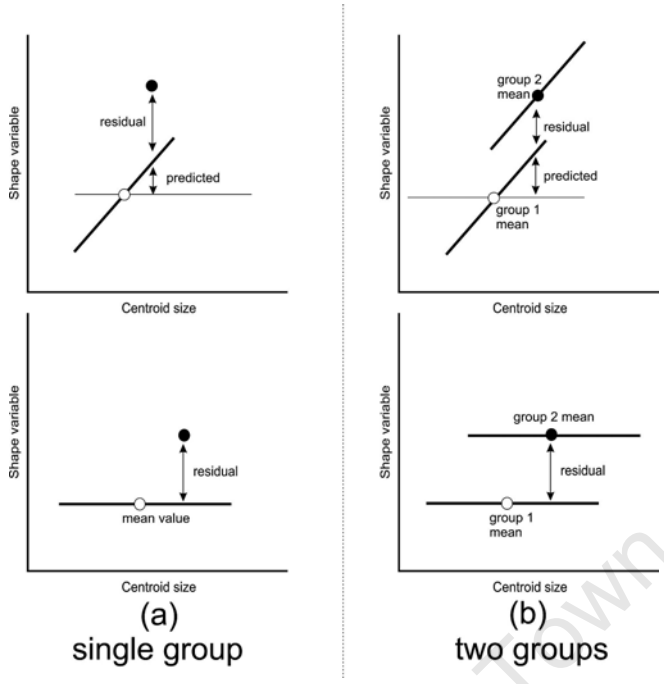

Figure 5.9: (a) Regression of single shape data point variable (black dot) on centroid size for a single group. The deviation from the sample average (hollow dot) in the direction of the shape variable (vertical; the dependent variable) is divided into a predicted and a residual component. The predicted component can be computed from the slope of the regression line (bold oblique line) and the deviation of the data point from the mean in the direction of centroid size (horizontal; the independent variable). The residual is the difference of the total shape deviation of the data point from the mean and the predicted component. (b) In the two group case, if the regression slopes in the groups are the same, a pooled within-group regression can be performed. This method uses the regression slopes within samples to separate the predicted and residual components of variation in the dependent variables.

the two groups and would provide some insight into the development of the FAS facial phenotype with age.

5.5.2 Results

There was a high correlation between centroid size and age when all the subjects were combined. Age explained 64% of the variation in centroid size ($p \approx 0$) for the permutation test against the null hypothesis of independence at the 95% confidence level. However, in a plot of centroid size vs. age (see Figure 5.10), FAS individuals showed smaller centroid

sizes.

The results of a regression of shape on age are shown in Table 5.4 for all ages combined and for the two age subgroups of five and twelve-year olds, separately. Age significantly (p=0.05) explained 6% of the shape variation for the combined subjects. For the five-year old subgroup, age did not significantly explain shape variability (p=0.31). After correction for age however, centroid size predicted 21% of the resultant shape variation (p=0.07) in the five-year old group. There was thus some weak evidence against the null hypothesis of independence between centroid size and age-corrected shape variables for the five-year old sub-group. In the twelve-year old sub-group, the results of the permutation tests for the null hypothesis of independence between age and shape and between centroid size and age-corrected shape were not significant.

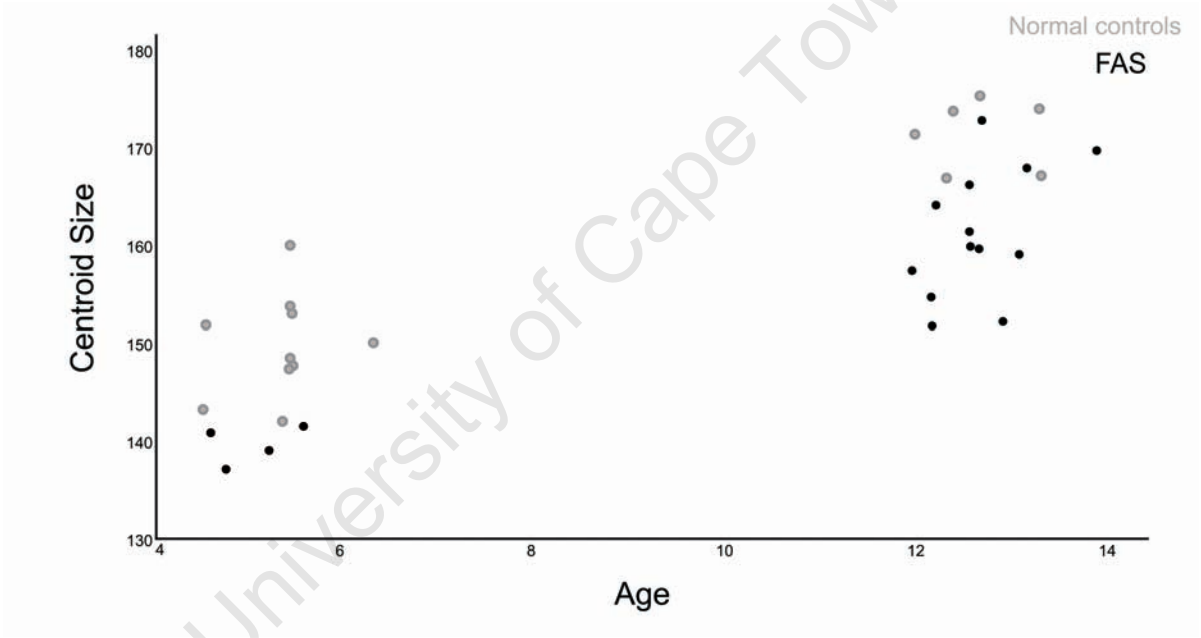

Figure 5.10: Comparison of centroid size in the FAS and the normal groups at different ages.

Shape differences were observed between the mean FAS and control shapes in five-year olds (Figure 5.11). The shape difference graph illustrates the shape change from the normal mean shape in the sample towards the FAS shape, by a scale factor. The scale factor chosen was 0.2, as it allowed for complete visualisation of shape differences while maintaining a biologically meaningful graphic.

Table 5.4: The results of a permutation test for independence after regression of shape on age and regression of the result on centroid size for all subjects combined.

		Regression of shape onto age	Regression of residuals of shape on age onto size
All ages combined	% predicted	5.98	1.76
	p-values	0.05	0.77
Five-year olds	% predicted	7.95	20.6
	p-values	0.31	0.07
Twelve-year olds	% predicted	5.28	4.29
	p-values	0.41	0.55

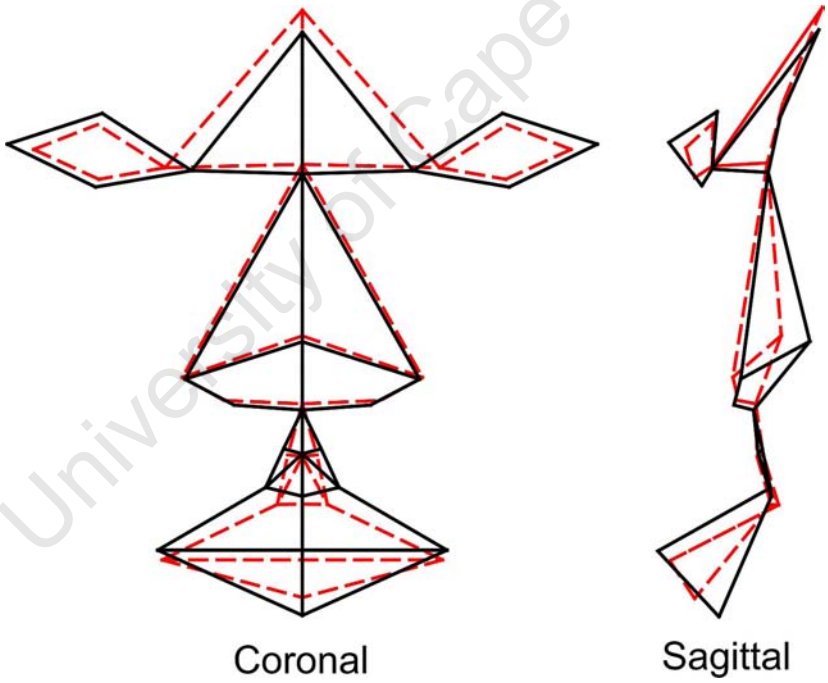

Figure 5.11: The shape difference between FAS and normal five-year old individuals given by the discriminant function analysis model after age and size correction. The solid lined black configuration is the normal mean shape and the dashed red line is the FAS mean shape. The scale factor is 0.2.

The glabella is more superiorly positioned in the FAS group. There is no great difference in shape in the midfacial region as viewed coronally, but the sagittal plane reveals a great ventro-dorsal retraction of the pronasale in the FAS mean shape. In the coronal view, the upper lip of the FAS group is also thinner and the philtrum length is much longer. The FAS lower lip is much thinner than the control lower lip. The difference in mean shapes is significant ($T^2 = 4289.35$, $p = 0.02$, see Table 5.5) and the discriminant function analysis cross-validation classification accuracy (Table 5.6) for the five-year olds reveals a sensitivity of 100% and a specificity of 90.9% (overall classification accuracy of 95.5%).

Table 5.5: Mean shape difference statistics, after correction for age and size, comparing the difference between FAS and normal mean facial configurations using 1) five year olds only, 2) twelve year olds only and 3) all subjects

	Five-year olds	Twelve-year olds	All subjects
T^2	4289.35	1848.49	22172.23
p-value (parametric)	0.02	0.07	0.02

Mean shape differences between FAS and control individuals at twelve years of age (see Figure 5.12) show, in the coronal view, reduced palpebral fissure lengths. There is very little difference in the position of the endocanthi in the two groups. The exocanthi in the FAS group, however, are more medially positioned accounting for the small palpebral fissures. There is also an up-slanting of the eyes for the FAS group. In the twelve-year old FAS sample, individuals have slightly thinner upper lips. The sagittal plane shows an upturning of the nose for control individuals, in contrast with the results for the combined analysis and that for five-year olds. The difference in mean shape for the twelve-year olds is weakly significant ($T^2 = 1848.49$, $p = 0.07$, Table 5.5). The cross-validation classification results of the discriminant function analysis show a FAS identification sensitivity of 76.9 % and a specificity of 83.3 % (overall classification accuracy of 80.1 %) (Table 5.6).

For all the subjects combined, the FAS mean shape seems to have smaller palpebral fissures. The endocanthi between the two mean shapes are almost in the same position but there is a great lateral difference in position of exocanthi between normal controls and FAS individuals, giving the FAS group smaller palpebral fissure lengths. The FAS mean shape also has a vertically longer nasal bridge length than the normal control mean shape.

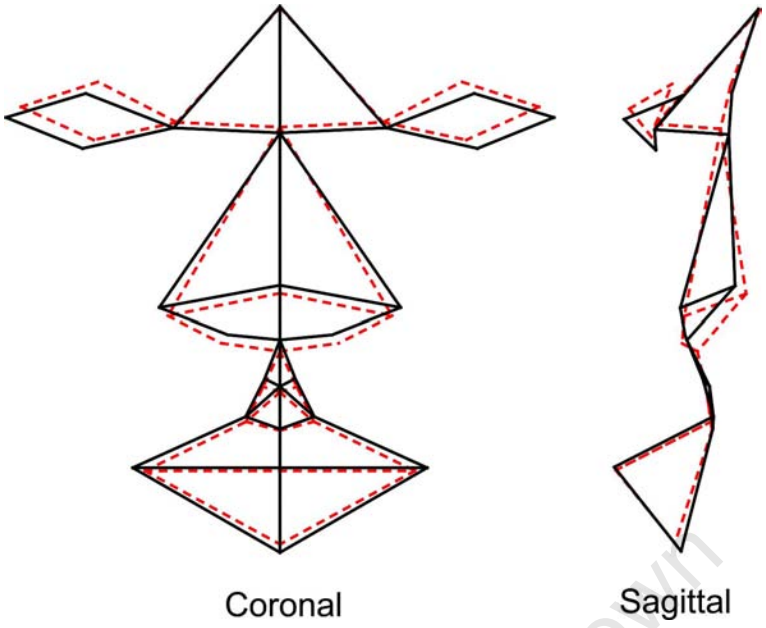

Figure 5.12: The shape difference between FAS and normal twelve-year old individuals given by the discriminant function analysis model after age and size correction. The solid lined black configuration is the normal mean shape and the dashed red line is the FAS mean shape. The scale factor is 0.2.

To complete the difference in shape in the coronal view, the FAS mean shape has a smaller upper lip and up-slanting eyes compared to the control mean shape. The difference in mean shape for the all subjects combined is significant (T^2 = 22172.23, p=0.02, Table 5.5). The cross-validation classification results of the discriminant function analysis show a FAS identification sensitivity of 88.2% and a specificity of 94.1 % (overall classification accuracy of 91.2 %) (Table 5.6).

The regression of shape onto age for the FAS group reveals that age predicts 43% of shape variability, although this is not significant (Table 5.7). In addition, the results of the subsequent regression of centroid size onto age-corrected shape variables show that centroid size predicts 15% of shape variability; this result is not statistically significant.

The results are similar for the control group; a non-significant prediction of age on shape variability of 47% and a non-significant prediction of centroid size on age-corrected shape of 7%. There are however, statistically significant differences in mean shape between the five and twelve-year olds for both the FAS and the control groups (see Table 5.8). The shape differences between the two age categories are shown in Figure 5.14. The shape

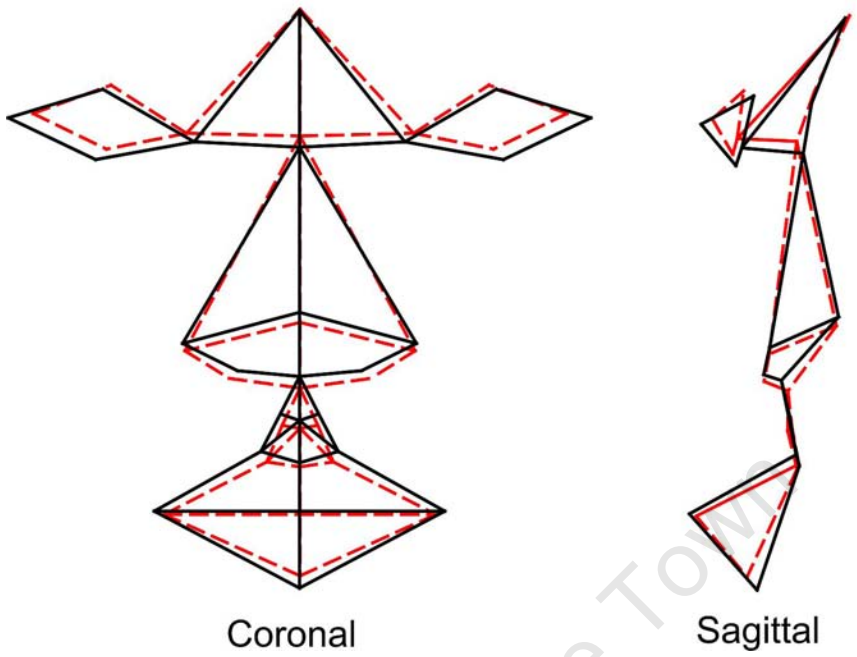

Figure 5.13: The shape difference between all FAS and normal individuals given by the discriminant function analysis model after age and size correction. The solid lined black configuration is the normal mean shape and the dashed red line is the FAS mean shape. The scale factor is 0.2.

Table 5.6: The cross-validation confusion matrix for classification using discriminant function analysis.

	Confusion Matrix						
	True group	FAS	Allocated to Normal	Total	Sensitivity %	Specificity %	Accuracy %
All ages combined	FAS	15	2	17	88.2	94.1	91.2
	Normal	1	16	17			
Five-year olds	FAS	4	0	4	100	90.9	95.5
	Normal	1	10	1			
Twelve-year olds	FAS	10	3	13	76.9	83.3	80.1
	Normal	1	5	6			

Table 5.7: The results of a permutation test for independence after regression of shape on age and regression of the result on centroid size for all subjects combined.

		Regression of shape onto age	Regression of residuals of shape on age onto size
FAS	% predicted	43.37	15.39
	p-values	0.25	0.17
Normal controls	% predicted	46.92	7.23
	p-values	0.19	0.23

Table 5.8: Mean shape difference statistics comparing 1) the FAS face at five with the FAS face at twelve and 2) the normal control face at five with the normal control face at twelve

	FAS	Normal controls
T^2	18042.14	5326.49
p-value (parametric)	0.01	0.02

difference between the five-year old FAS sample and the twelve-year old FAS sample is very pronounced. In particular, the glabella in the five-year old FAS mean shape is much more superiorly positioned than in the twelve-year old FAS mean shape. In addition there is a marked difference in the position of the nasion and pronasale between the two age groups when viewing the shapes sagittally. In contrast to the FAS group, the differences in shape between the five-year old and twelve-year old control groups are much less pronounced.

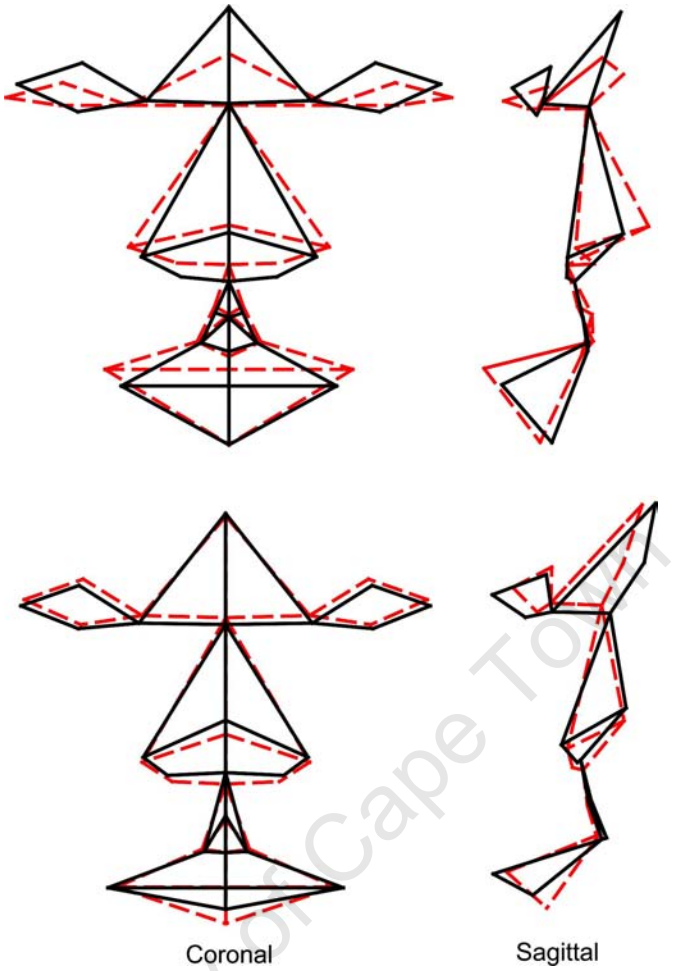

Figure 5.14: The shape difference between FAS five year olds and FAS twelve year olds (top) and control five year olds and control twelve year olds (bottom). The black solid lined is the five year old mean shape and the red dashed line is the twelve year old mean shape. The scale factor used was 0.05.

5.6 Chapter discussion and original contribution

Multivariate statistical procedures applied to distances, angles and distance ratios have previously been employed in the assessment of facial anomalies associated with FAS (Astley & Clarren, 1995, 1996; Moore *et al.*, 2001, 2002, 2007; Naidoo *et al.*, 2006). Geometric morphometric (in particular, here, facial shape analysis) methods retain all geometric information while simultaneously providing a statistical platform to investigate biological organisms in a quantitative way. In addition, the results of the analyses can be readily related back to the physical space of the specimens and graphically viewed. Previously,

it was shown that facial shape analysis could play an important role in characterizing the FAS facial phenotype (Mutsvangwa & Douglas, 2007). Here, the methodology has been refined through explicit considerations of some factors that affect shape variability. These include symmetry, age and size. In addition, classification of subjects has been introduced.

Discriminant analysis of the Procrustes residuals did not reveal any statistically significant differences in the mean shapes of subjects for any of the age groups and for the subjects combined. This suggests that the FAS facial anomalies in this sample were too subtle to be detected using the mean facial shapes from the GPA alone or that the within-group variation is large compared to the between-group variation. This illustrates one of the shortcomings of geometric morphometrics, namely that a large sample is needed to mitigate against the effects of within-group variation and reveal the sometimes subtle differences that exist between a FAS face and a normal face.

With the first geometric morphometric approach, it was possible to isolate discriminating shape variation from the sample regardless of the initial result of no significant difference in mean shapes. In particular, for the five-year olds, large shape differences were observed using the method of feature selection of principal components. The principal components selected accounted for a considerable amount of shape variation, 22.7% in total. The classification results for the first approach showed a high level of accuracy with an average of 85% for all the pattern recognition algorithms. The second approach using regression for extrinsic factor correction showed even higher classification accuracy at 95.5% for the five-year old age group. In addition, the mean facial differences between FAS and control subjects in the five-year old group, using the second approach, was statistically significant. The facial differences between FAS and controls in this age group followed those reported in the literature. Table 5.15 summarizes the FAS facial features for all ages and also compares them for the two approaches. The FAS shape presents small palpebral fissures and thin upper lips, both of which have been associated with the condition. Moore *et al.* (2007) reported short philtrum length and reduced inner canthal length as facial features that discriminate individuals with FAS in a sample of mixed ancestry subjects with a mean age of five years. The results here show an increased philtrum length and increased inner canthal distance in FAS subjects in a different sample of the same population, which corresponds with the results we reported previously for seven-year olds (Mutsvangwa & Douglas, 2007). The results presented here agree with those of Moore *et al.* (2007) in the observation that palpebral fissure lengths are shorter as a result of shorter outer canthal distance rather than because of changes in the inner

canthal length for FAS subjects. In previous work we found significant hypoplasia of the midface (Mutsvangwa & Douglas, 2007). Here, midface hypoplasia is the most significant anomaly in the FAS five-year old group, visually.

The popular research opinion is that the FAS facial shape anomalies diminish with age (Astley & Clarren, 1995; Moore *et al.*, 2007). In both approaches, the classification results for the twelve-year old categories was poorer than for the five-year olds. In addition, the shape differences between normal and FAS subjects were much more subtle for the twelve-year olds. Using the first approach, the principal components chosen accounted for much less of the total shape variation than for the five-year old group (a total variation of 5.6% for the two selected principal components). In the second approach, the mean facial differences between FAS and control subjects in the twelve-year old group were only weakly significant. It is important to note, however, that the selection of the older age group possibly included individuals at very different stages of their adolescent growth spurt accounting for greater variation in this group. This limitation could be addressed through a longitudinal study approach to the analysis. This was not possible within the time frame of the project.

Moore *et al.* (2007) concluded that some of the differences in the FAS facial phenotype across ethnicities are age-related. They included age as one of the discriminating features between FAS and control subjects for their mixed ancestry population. In our population of the same ethnicity, there are greater differences in facial shape between FAS individuals at different age groups than for control individuals. This supports the notion that FAS facial anomalies disappear with age as there seems to be a greater difference in shape change in this group from a uniquely FAS-like face at five to a facial shape less distinct from normal by twelve years of age. However, some differences in the facial shape of control and FAS subjects persisted, as evidenced by the diminished nasal protrusion and thinner upper lips, anomalies which were common for the FAS shape in both analytical approaches.

The results for all subjects combined using the first approach showed a low classification accuracy. The principal components chosen accounted for a considerable amount of shape variation (25.6% of total shape variation), however, the first principal component was selected. In the absence of any size correction it is thought that the first principal component represents size. Thus the selection of the most significant principal component as a discriminator between FAS and control subjects might have been due to the uneven matching between the two groups in terms of age (see Table 5.1). However, the the regression of principal component 1 on centroid size showed that size only predicted

12.1% of the shape variation in this principal component. The other principal component selected by the feature selection procedure was principal component 15, which accounted for a very small percentage of shape variation and revealed subtle differences in shape between FAS subjects and normal controls.

The second approach showed a high classification accuracy of 91.2% for the combined group, a substantial improvement on the classification results of the first approach. The FAS facial shape in both approaches for all subjects combined included shorter palpebral fissures, up-slanting of the eyes, a vertically longer nasal bridge and slighlty thinner upper lips.

From the results of the second approach, age was significantly correlated with centroid size when all subjects were combined. If centroid size could be used here as a proxy for facial size (an acceptable assumption because the landmarks are distributed around the face), then this result is expected, as older subjects are expected to have larger faces. FAS subjects had smaller facial size compared to controls of the same age group. In younger subjects, this difference in facial size was greater than in the older sample. Thus, from a size perspective, the results suggest that FAS faces are generally smaller at younger ages compared to control faces, and, as individuals grow, the differences in head size diminish.

The smoothness of the philtrum furrow, another facial anomaly associated with FAS, was visually missing in the FAS facial shapes for all analyses. A possible explanation for this would be the poor definition of the philtrum shape. This feature is hard to define using landmarks only and highlights a critical short-coming of the landmark-based approach presented here. While it is relatively easy to include a facial feature defined by landmarks located on points of biological meaning or extreme curvature, it is difficult if not impossible to use a 2D image to locate or isolate features that consist of soft tissue alone.

Fang *et al.* (2008) developed a novel approach which automatically detects facial features using 3D facial images from a laser scanner. Computer graphics, machine learning and pattern recognition methods were employed to determine a list of features that best differentiate subjects with FAS and normal individuals. Their methodology, while technically complex, presents a shift in FAS diagnosis in that it makes no prior assumptions about the selection of facial features to be measured. Their method, which was applied to two different ethnic populations (50 FAS and 32 control Finish Caucasian; 36 FAS and 31 control mixed ancestry subjects from the Cape region in South Africa) could correctly classify: the Finish sample with 88.2% sensitivity and 100% specificity, the mixed ancestry sample with 91.7% sensitivity and 90% specificity; and the combined group with

82.8% sensitivity and 76.2% specificity. The analysis here, using the second approach, had comparable results. However, it relied on the use of within-group regression in all the regression steps. This, by design, maximizes the separation of groups. While here serving to achieve the greatest visual facial differences between FAS and control subjects, it reduces the significance of the classification model as apriori information about the grouping of a subject is included in the regression before classification. Fang *et al.* (2008) did not report on the visual differences in facial features between FAS and control subjects in their analysis. Rather, their methodology selected different facial features depending on which pattern classification technique they used.

Both approaches presented here have advantages and disadvantages when compared to each other. The first approach makes no a priori assumptions about group membership structure thus providing a true classification accuracy compared to the second approach. The first method, however, does not directly reveal known extrinsic factors that might be responsible for some of the shape variation as in the second approach. In addition, it may be too much of a mathematical abstraction of the underlying biological processes and thus conclusions drawn using this method need to be drawn with care. Choosing a principal component from feature selection, based on a distance measure in data space as was done with the branch-and-bound feature selection algorithm, is an example of that abstraction. The second method offers a more biologically fundamental comparison of specimens although some assumptions of the structure of the biological data have to be made.

A significant limitation of both approaches was the unbalanced sample used in the analysis. The study sample was not sex or age group matched. The effects of sexual dimorphism on the FAS facial shape, however, were explored in a previous study and were found not to be significant (Mutsvangwa & Douglas, 2007). The sample was derived from a mixed ancestry population allowing for the comparison of our results with those of other studies in the same ethnic group (Fang *et al.*, 2008; Moore *et al.*, 2007). The small study sample is a limitation; however, the narrow age bands studied allowed insights into shape differences at different ages. In addition, this type of 3D shape analysis is a new tool in assessing the FAS facial phenotype and we have presented results that hold promise for objective diagnosis of FAS.

Age category	Approach one	Approach two	Similarities of two approaches
Five year olds	• Large reduction in the size of the eye orbits. • Larger inner canthal and smaller outer canthal distances for the FAS mean resulting in smaller palpebral fissure lengths compared to the control mean. • The glabella also changes position dramatically from a more inferior to a more superior position. • A great ventro-dorsal retraction of the pronasale and upturning of the nose. • Thinning of the upper lip, increase in philtrum length.	• In general there is a great reduction of the eye orbits. • Larger inner canthal and smaller outer canthal distances for the FAS mean resulting in smaller palpebral fissure lengths compared to the control mean. • The glabella is more superiorly positioned in the FAS group. • A great ventro-dorsal retraction of the pronasale and upturning of the nose in the FAS face. • A significantly thinner upper lip and longer philtrum length.	• All shape changes occur in both approaches.
Twelve year olds	• A ventro-dorsal change in the relative position of the nasion and seillion with glabella changing in the opposite direction. • Evidence of the nasal protrusion diminishing. • Slightly thinner upper lips.	• Reduced palpebral fissure lengths. • There is also an up-slanting of the eyes for the FAS group. • Slightly thinner upper lips. • Upturning of the nose for control individuals,.	• Evidence of the nasal protrusion diminishing. • Slightly thinner upper lips.
All ages combined	• Shorter palpebral fissures and down slanting of the eyes. • Large dorso-ventral and infero-superior movement of the glabella and nasion and seillion, although the nasion shows a greater change in relative position. • Shorter midfacial length from the seillion to the labiale superius. • Thinner upper- and lower-lips.	• Shorter palpebral fissures and down slanting of the eyes. • Vertically longer nasal bridge length than the normal control mean shape. • Smaller upper lip and up-slanting eyes.	• All shape changes similar except for the relative change in the glabella, nasion and seillion.

Figure 5.15: A comparison of the FAS facial presentation for both approaches.

Chapter 6

Overview

This thesis reports on the design of a stereo-photogrammetric imaging tool, on the testing of the imaging tool and on methods that may be employed to analyse data obtained from the tool.

The conflicting goals in designing such an anthropometric imaging tool were low cost vs. quality of information. The system had to provide clinically accurate measurements of the human face which are comparable to those produced by other imaging systems. The research question at hand becomes the determining factor in balancing cost and accuracy.

In most studies reported in the literature, the characterization of the facial phenotype associated with FAS has relied on anthropometric measurements of facial features associated with the syndrome, by way of linear distance measurements obtained directly from photographs and Likert scales. Data for such assessment is directly accessible from images obtained with the imaging tool presented here. One advantage of the imaging tool over traditional methods of obtaining linear distances is that the data can also be used in geometric morphometric analysis, which allows for a comparison and averaging of facial shapes in terms of the relative 3D positions of a set of landmarks for shape variation.

The literature abounds with reports of sophisticated commercial imaging technologies employed in surgical planning, post-surgical assessment, assessment of longitudinal change in morphology and quantification of syndromic clinical features (Fang *et al.*, 2008; Ghoddousi *et al.*, 2007; Gwilliam *et al.*, 2006; Hajeer *et al.*, 2002; Hammond *et al.*, 2004; Hennessy *et al.*, 2002, 2005; Honrado & Larrabee, 2004; Majid *et al.*, 2005; Weinberg *et al.*, 2004). An advantage of these systems over the imaging tool presented here is that surface information of the face is easily obtainable and can be included in the geometric morphometric analysis. In addition, the identification of landmarks is easier on

one composite 3D image than on three 2D images. However, these commercial systems are generally prohibitively expensive, costing tens of thousands of dollars per system (http://www.thinglab.co.uk).

The system presented here showed a high degree of reliability and precision, when compared with direct anthropometry using an annotated doll. The results were comparable with those reported in the literature using the same protocol of employing an annotated doll in assessing reliability and precision (Kovacs *et al.*, 2006; Weinberg *et al.*, 2004). The use of a doll eliminated pose and other subject-related errors, allowing for an assessment of the system alone. In a controlled laboratory situation, using an annotated inanimate doll, the information that can be obtained using the imaging tool is as reliable as, and more precise than, that obtained from direct anthropometry. In addition, the pose assessment reliability results using human subjects indicated, in most instances, high reliability.

Mixed results were obtained when eye measurements, obtained using the imaging tool, were compared with those taken by a dysmorphologist on human subjects in the field. The priority during the field work, however, was not to assess the comparability of the imaging modalities. The study comparing the Vectra 3D and the stereo-photogrammetry tool had the same limitations.

The portability of the imaging tool has allowed for the imaging of subjects in remote areas of the Northern Cape in South Africa. In the field, the imaging tool has proven to be an easy-to-use, subject-friendly device. A large database of images is currently being developed, providing a permanent record of subjects and also data for future analytical studies.

Two geometric morphometric approaches were presented that utilized the 3D landmark coordinate data from images taken using the imaging tool. One approach used generalised Procrustes analysis and principal component analysis to characterize the FAS facial phenotype in mixed race subjects. The other approach used regression for correction of age- and size-related shape variability before using discriminant function analysis to characterize the FAS facial phenotype in the same subjects. The first approach makes no a-priori assumptions about group membership structure thus providing a more convincing classification accuracy compared to the second approach. The first method, however, does not directly reveal known extrinsic factors that might be responsible for some of the shape variation as does the second approach. The second method thus offers a more biologically fundamental comparison of specimens although some assumptions of the structure of the biological data have to be made. The shape differences obtained in all the analyses largely

follow those reported in the literature. The FAS mean shapes present small palpebral fissures and thin upper lips, both of which have been associated with the condition. In both approaches, the facial shape differences between FAS and normal controls were more pronounced in the five year-old sample than in the twelve year-old sample. This validates the popular research opinion that the FAS facial anomalies diminish with age. In addition, the classification results showed higher accuracy in classifying the five year-olds than the twelve year-olds.

Traditionally, the gestalt method depended on the combined presentation of facial anomalies on the face. Thus, no one feature dominated in the diagnosis. Currently, the diagnostic criteria for the FAS facial phenotype assess each individual anomaly on the face and consider the number of anomalies rather than their resulting combined presentation. This thesis employs a novel approach to assessment of the FAS facial phenotype which treats the face as a whole unit and not the sum of individual features. In this way, facial shape analysis is more similar to the gestalt method, but, with the added benefit of being a quantitative approach rather than a qualitative one. As the database of facial images of both FAS and control subjects grows, it is hoped that the limitation of a small study sample will eventually be overcome.

6.1 Conclusions and recommendations

The system performed well in precision and reliability tests involving an inanimate object. Its reliability when used on children was lower, but given the limitations of imaging children, particularly infants, these results may be considered acceptable, especially as the system allows for repeated capture of photographs when the subject's pose is not suitable.

Future work should compare the stereo-photogrammetric system and the Vectra 3D using an inanimate object as well as human subjects under laboratory conditions. Future work could also include the exploration of structured-light stereo-photogrammetry methods to improve the imaging tool. With the addition of a projector that illuminates the subject with a pattern of light, and the three currently available cameras, such a system could, in addition to providing a 3D surface of the face, also eliminate the need for a calibration frame during imaging. This would improve the user-friendliness of the imaging tool, making it easier to obtain images of infants.

Using the Procrustes aligned facial shapes as data for cluster analysis is another avenue for future research. Cluster analysis is an explorative data analysis tool that aims to

ascribe group membership based on similarity; similar objects cluster closely together, and far away from dissimilar objects. The Procrustes distance obtained after alignment of facial shapes can readily be used as a similarity measure. Such clustering will provide a classification of subjects which will not need a-priori diagnostic information.

In conclusion, this thesis has demonstrated that stereo-photogrammetrically derived 3D facial landmarks and geometric morphometrics can be useful in characterizing the FAS facial phenotype and in visualizing and classifying facial shape differences. The imaging tool, while not able to produce 3D surfaces, provides useful information in the form of 3D coordinates. In addition, the tool is inexpensive, both to construct and to use, and its portability makes it a practical way to obtain data for anthropometric studies.

The methods presented here do not seek to remove the role of a trained dysmorphologist in FAS diagnosis, but instead to investigate the merits of using alternative methods that may aid a dysmorphologist during large-scale screening programmes for FAS. This is of particular importance to South Africa which has many competing health issues including tuberculosis, sexually transmitted diseases, particularly, acquired immunodeficiency syndrome (Viljoen *et al.*, 2003), while at the same time having the highest prevalence of FAS in the world.

Appendix A

Table A.1: Principal components showing percentage of shape variation each principal component represents for for each of the analyses

PC	Five year-olds	Twelve year-olds	All subjects
	Variance explained %)	Variance explained (%)	Variance explained (%)
1	29.748	29.665	24.627
2	15.782	18.224	16.014
3	13.650	11.534	10.375
4	9.059	9.173	7.647
5	6.870	8.198	6.334
6	6.383	4.883	6.188
7	5.086	3.910	4.999
8	4.533	3.116	4.206
9	2.957	2.811	3.924
10	2.463	2.026	2.809
11	1.421	1.713	2.451
12	0.817	1.257	1.836
13	0.660	1.038	1.500
14	0.571	0.766	1.329
15		0.674	0.971
16		0.457	0.899
17		0.351	0.771
18		0.205	0.644
19			0.464
20			0.456
21			0.338
22			0.279
23			0.261
24			0.200
25			0.150
26			0.105
27			0.079
28			0.049
29			0.037
30			0.023
31			0.016
32			0.011
33			0.006

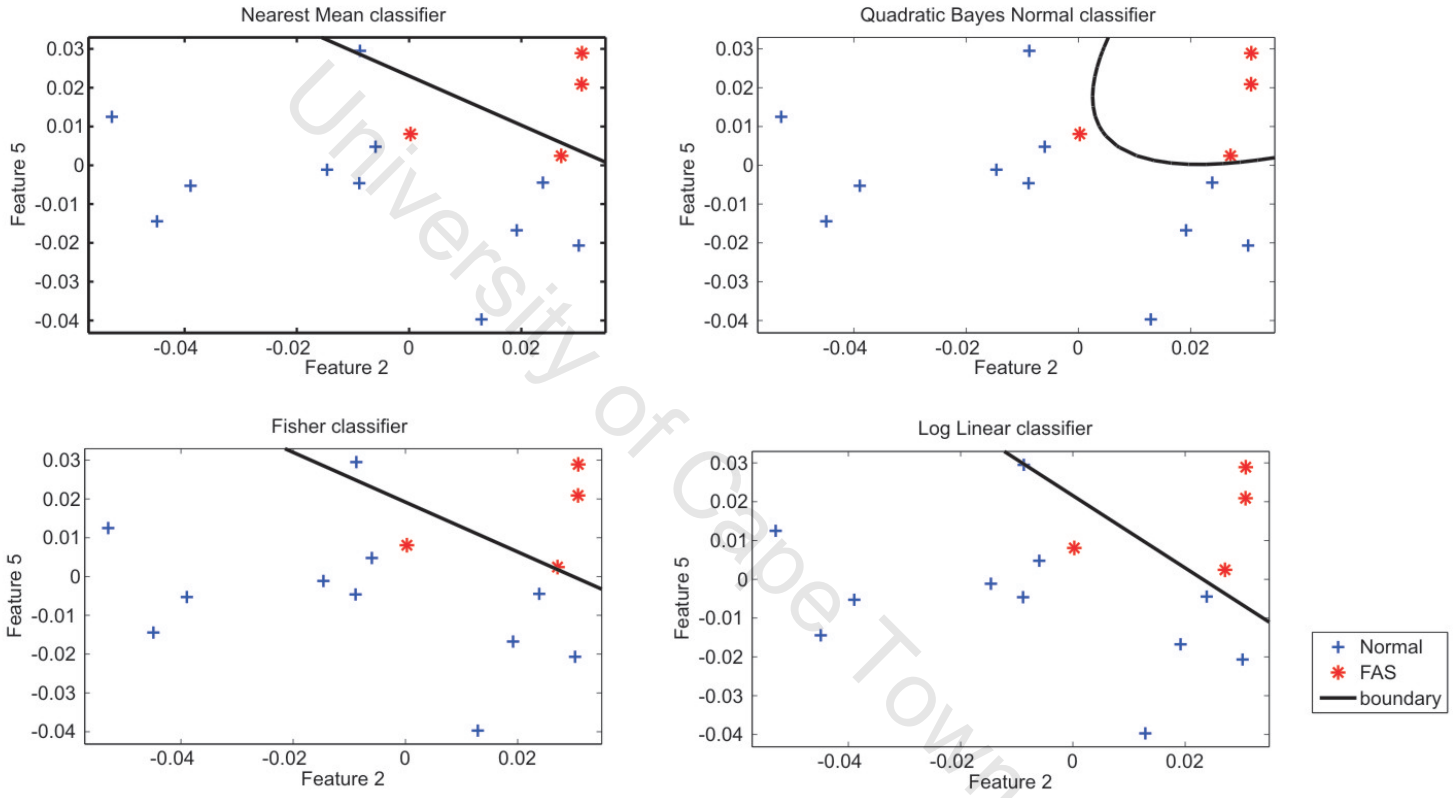

Figure A.1: Feature selection and classification of five year old subjects for one of the trials in leave-one-out cross-validation. The most discriminating principal components are PC2 and PC5.

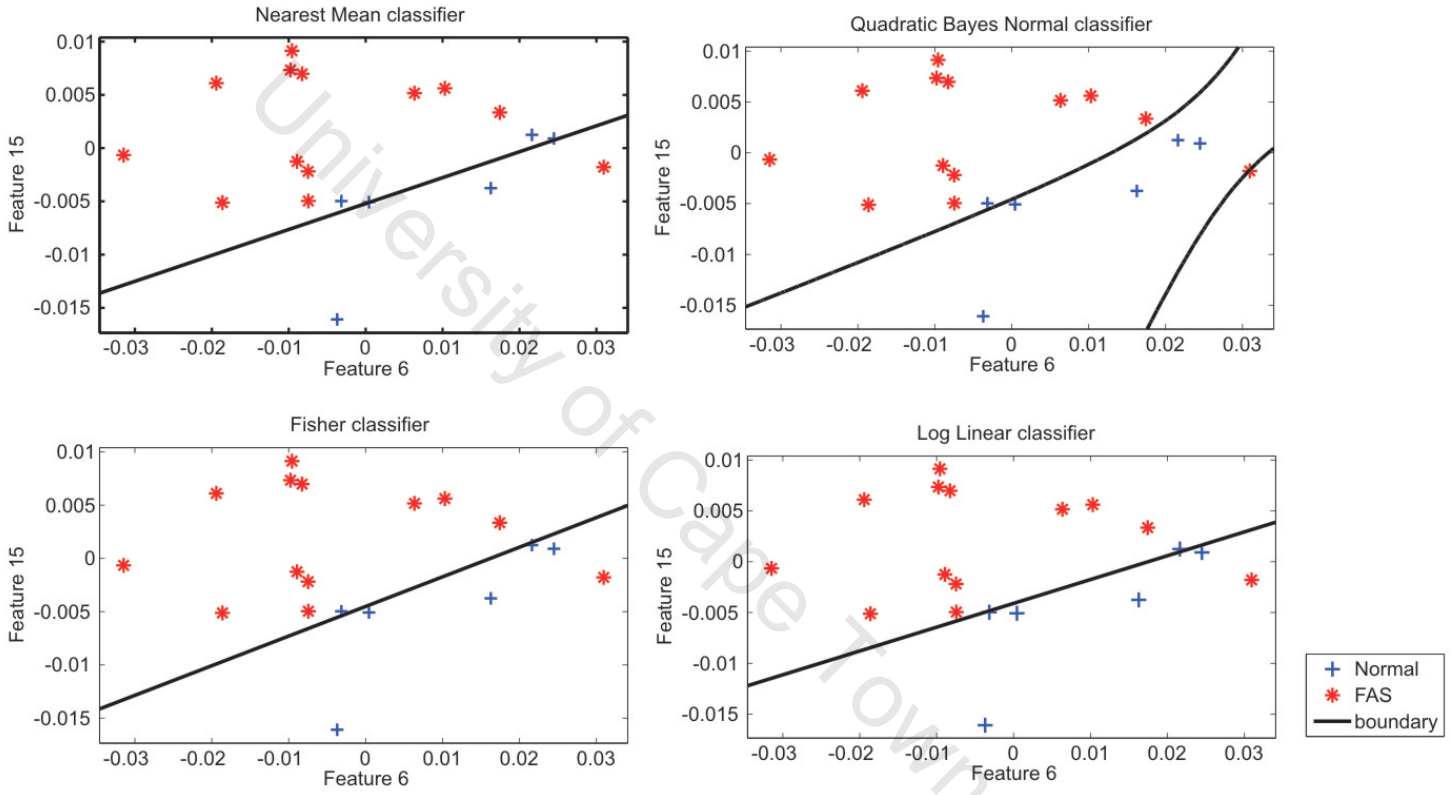

Figure A.2: Feature selection and classification of twelve year old subjects for one of the trials in leave-one-out cross-validation. The most discriminating principal components are PC6 and PC15.

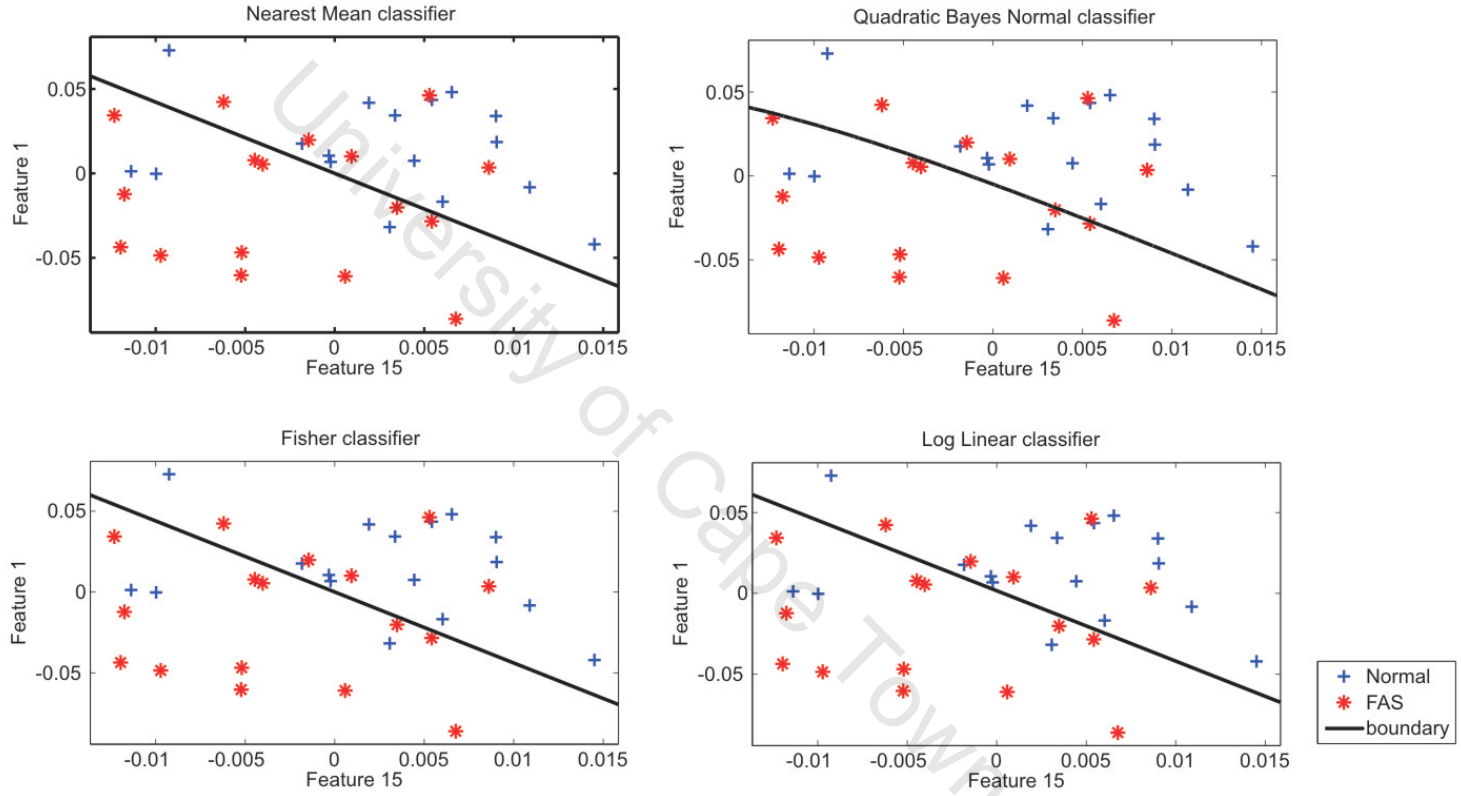

Figure A.3: Feature selection and classification of all subjects for one of the trials in leave-one-out cross-validation. The most discriminating principal components are PC1 and PC15.

References

Abdel-Aziz, Y. & Karara, H. (1971). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Proceedings of the Symposium on Close-Range Photogrammetry*, 1–18. One citation on page 26.

Aldridge, K., Boyadjiev, S.A., Capone, G.T., V, D. & Richtsmeier, J.T. (2005). Precision and error of three-dimensional phenotypic measures acquired from 3dmd photogrammetric images. *Am J Med Genet A*, **138**, 247–53. 2 citations on pages 45 and 64.

Astley, S.J. (2004). *Diagnostic Guide for Fetal Alcohol Spectrum Disorders: The 4-Digit Diagnostic Code*. University Publication Services, 3rd edn. One citation on page 8.

Astley, S.J. (2006). Comparison of the 4-digit diagnostic code and the hoyme diagnostic guidelines for fetal alcohol spectrum disorders. *Pediatrics*, **118**, 1532–1545. 3 citations on pages 5, 12, and 20.

Astley, S.J. & Clarren, S.K. (1995). A fetal alcohol syndrome screening tool. *Alcohol Clin Exp Res*, **19**, 1565–71. 13 citations on pages 1, 2, 5, 6, 7, 8, 9, 12, 15, 74, 75, 94, and 96.

Astley, S.J. & Clarren, S.K. (1996). A case definition and photographic screening tool for the facial phenotype of fetal alcohol syndrome. *J Pediatr*, **129**, 33–41. 7 citations on pages 2, 6, 8, 12, 15, 74, and 94.

Astley, S.J. & Clarren, S.K. (2000). Diagnosing the full spectrum of fetal alcohol-exposed individuals: introducing the 4-digit diagnostic code. *Alcohol Alcohol*, **35**, 400–10, 0735-0414 Journal Article. 8 citations on pages 2, 5, 6, 8, 12, 13, 20, and 74.

Aung, S.C., Ngim, R.C.K. & Lee, S.T. (1995). Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements. *British Journal of Plastic Surgery*, **48**, 551–558. One citation on page 40.

Burd, L., Martsolf, J.T., Klug, M.G. & Kerbeshian, J. (2003). Diagnosis of fas: a comparison of the fetal alcohol syndrome diagnostic checklist and the institute of medicine criteria for fetal alcohol syndrome. *Neurotoxicol Teratol*, **25**, 719–24. 4 citations on pages 1, 5, 6, and 7.

Cangelosi, R. & Goriely, A. (2007). Component retention in principal component analysis with application to cdna microarray data. *Biology Direct*, **2**, 2. One citation on page 77.

Chong, A.K. & Mathieu, R. (2006). Near infrared photography for craniofacial anthropometric landmark measurement. *The Photogrammetric Record*, **21**, 16–28. 3 citations on pages 17, 36, and 64.

Choo, A.M.T. & Oxland, T.R. (2003). Improved rsa accuracy with dlt and balanced calibration marker distributions with an assessment of initial-calibration. *Journal of Biomechanics*, **36**, 259–264. One citation on page 33.

Chudley, A.E., Conry, J., Cook, J.L., Loock, C., Rosales, T. & LeBlanc, N. (2005). Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *CMAJ*, **172**, S1–21. One citation on page 6.

Clarren, S.K., Sampson, P.D., Larsen, J., Donnell, D.J., Barr, H.M., Bookstein, F.L., Martin, D.C. & Streissguth, A.P. (1987). Facial effects of fetal alcohol exposure: assessment by photographs and morphometric analysis. *Am J Med Genet*, **26**, 651–66. 7 citations on pages 1, 5, 6, 8, 11, 66, and 74.

Cox-Brinkman, J., Vedder, A., Hollak, C., Richfield, L., Mehta, A., Orteu, K., Wijburg, F. & Hammond, P. (2007). Three-dimensional face shape in fabry disease. *Eur J Hum Genet*, **15**, 535–542. One citation on page 54.

Croxford, J. & Viljoen, D. (1999). Alcohol consumption by pregnant women in the western cape. *S Afr Med J*, **89**, 962–5. One citation on page 7.

D'Apuzzo, N. (2002). Surface measurement and tracking of human body parts from multi-image video sequences. *ISPRS Journal of Photogrammetry and Remote Sensing,*, **56**, 360–375. One citation on page 17.

Douglas, T., Meintjes, E., Vaughan, C. & Viljoen, D. (2003a). Comparison of eye and lip measurements obtained using single and stereo-photogrammetry. In *Proceedings of the 25th Annual International Conference of the IEEE*, vol. 1, 545– 547. One citation on page 14.

Douglas, T.S. (2004). Image processing for craniofacial landmark identification and measurement: a review of photogrammetry and cephalometry. *Comput Med Imaging Graph*, **28**, 401–9. One citation on page 13.

Douglas, T.S. & Viljoen, D.L. (2006). Eye measurements in 7-year-old black south african children. *Annals of Human Biology*, **33**, 241–254. One citation on page 19.

Douglas, T.S., Martinez, F., Meintjes, E.M., Vaughan, C.L. & Viljoen, D.L. (2003b). Eye feature extraction for diagnosing the facial phenotype associated with fetal alcohol syndrome. *Med Biol Eng Comput*, **41**, 101–6. 4 citations on pages 13, 14, 38, and 62.

Douglas, T.S., Meintjes, E.M., Vaughan, C.L. & Viljoen, D.L. (2003c). Role of depth in eye distance measurements: comparison of single and stereo-photogrammetry. *American Journal of Human Biology*, **15**, 573–578. One citation on page 14.

Dryden, I.L. & Mardia, K.V. (1998). *Statistical Shape Analysis*. John Wiley & Sons, New York. 2 citations on pages 66 and 67.

Duin, R.P.W. (2000). Prtools version 3.0: A matlab toolbox for pattern recognition. 2 citations on pages 75 and 79.

Fang, S., McLaughlin, J., Fang, J., Huang, J., Autti-Ramo, I., Fagerlund, A., Jacobson, S., Robinson, L., Hoyme, H., Mattson, S., Riley, E., Zhou, F., Ward, R., Moore, E. & Foroud, T. (2008). Automated diagnosis of fetal alcohol syndrome using 3d facial image analysis. *Orthodontics & Craniofacial Research*, **11**, 162–171. 7 citations on pages 8, 17, 18, 19, 97, 98, and 100.

Farkas, L. (1994). *Anthropometry of the head and face*. Raven Press, New York. 4 citations on pages 8, 11, 13, and 40.

Frank, A. (2002). *A new branch and bound feature selection algorithm*. Master's thesis, Israel Institute of Technology. One citation on page 78.

Ghoddousi, H., Edler, R., Haers, P., Wertheim, D. & Greenhill, D. (2007). Comparison of three methods of facial measurement. *International Journal of Oral and Maxillofacial Surgery*, **36**, 250–258. 5 citations on pages 17, 40, 57, 64, and 100.

Grobbelaar, R. & Douglas, T.S. (2007). Stereo image matching for facial feature measurement to aid in fetal alcohol syndrome screening. *Medical Engineering & Physics,*, **29**, 459–464. 2 citations on pages 13 and 16.

Gwilliam, J.R., Cunningham, S.J. & Hutton, T. (2006). Reproducibility of soft tissue landmarks on three-dimensional facial scans. *The European Journal of Orthodontics*, **28**, 408–415. 4 citations on pages 17, 42, 64, and 100.

Hajeer, M.Y., Ayoub, A.F., Millett, D.T., Bock, M. & Siebert, J.P. (2002). Three-dimensional imaging in orthognathic surgery: the clinical application of a new method. *Int J Adult Orthodon Orthognath Surg*, **17**, 318–30. 2 citations on pages 17 and 100.

Halazonetis, D.J. (2004). Morphometrics for cephalometric diagnosis. *Am J Orthod Dentofacial Orthop*, **125**, 571–81. 5 citations on pages 16, 68, 69, 70, and 71.

Hall, J.G., Froster-Iskenius, U.G. & Allanson, J.E. (1989). *Handbook of normal physical measurements*. Oxford University Press, Oxford; New York. One citation on page 19.

Hammond, P. (2007). The use of 3D face shape modelling in dysmorphology. *Arch Dis Child*, **92**, 1120–1126. One citation on page 54.

Hammond, P., Hutton, T.J., Allanson, J.E., Campbell, L.E., Hennekam, R.C., Holden, S., Patton, M.A., Shaw, A., Temple, I.K., Trotter, M., Murphy, K.C. & Winter, R.M. (2004). 3d analysis of facial morphology. *Am J Med Genet A*, **126**, 339–48. 4 citations on pages 16, 17, 37, and 100.

Hennessy, R.J., Kinsella, A. & Waddington, J.L. (2002). 3d laser surface scanning and geometric morphometric analysis of craniofacial shape as an index of cerebrocraniofacial morphogenesis: initial application to sexual dimorphism. *Biological Psychiatry,*, **51**, 507–514. 4 citations on pages 17, 65, 70, and 100.

Hennessy, R.J., McLearie, S., Kinsella, A. & Waddington, J.L. (2005). Facial surface analysis by 3d laser scanning and geometric morphometrics in relation to sexual dimorphism in cerebral-craniofacial morphogenesis and cognitive function. *Journal of anatomy*, **207**, 283–295. 3 citations on pages 17, 18, and 100.

Hill, T. & Lewicki, P. (2007). *Statistics methods and applications*. StatSoft, Tulsa,OK. One citation on page 77.

Honrado, C.P. & Larrabee, W.F.J. (2004). Update in three-dimensional imaging in facial plastic surgery. *Curr Opin Otolaryngol Head Neck Surg*, **12**, 327–31. 2 citations on pages 37 and 100.

Hoyme, H.E., May, P.A., Kalberg, W.O., Kodituwakku, P., Gossage, J.P., Trujillo, P.M., Buckley, D.G., Miller, J.H., Aragon, A.S., Khaole, N., Viljoen, D.L., Jones, K.L. & Robinson, L.K. (2005). A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 institute of medicine criteria. *Pediatrics*, **115**, 39–47. 3 citations on pages 1, 20, and 74.

Huang, J., Jain, A., Fang, S. & Riley, E.P. (2005). Using facial images to diagnose fetal alcohol syndrome (fas). In *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05)*, 66–71, IEEE Computer Society. 3 citations on pages 8, 13, and 17.

Hunter, A.G. (2002). Medical genetics: 2. the diagnostic approach to the child with dysmorphic signs. *CMAJ*, **167**, 367–72. One citation on page 5.

Jamison, P.L. & Ward, R.E. (1993). Brief communication: measurement size, precision, and reliability in craniofacial anthropometry: bigger is better. *Am J Phys Anthropol*, **90**, 495–500. 3 citations on pages 42, 59, and 61.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer, New York, 2nd edn. One citation on page 77.

Jones, K.L. (1986). Fetal alcohol syndrome. *Pediatr Rev*, **8**, 122–6. One citation on page 5.

Jones, K.L. & Smith, D.W. (1973). Recognition of the fetal alcohol syndrome in early infancy. *Lancet*, **2**, 999–1001. 2 citations on pages 5 and 6.

Klingenberg, C.P. (2008). *MorphoJ*. Faculty of Life Sciences, University of Manchester. 3 citations on pages 75, 77, and 85.

Klingenberg, C.P. & Monteiro, L.R. (2005). Distances and directions in multidimensional shape spaces: implications for morphometric applications. *Syst Biol*, **54**, 678–88. One citation on page 65.

Klingenberg, C.P., Barluenga, M. & Meyer, A. (2002). Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution*, **56**, 1909–1920. One citation on page 72.

Klug, M.G. & Burd, L. (2003). Fetal alcohol syndrome prevention: annual and cumulative cost savings. *Neurotoxicology and Teratology*, **25**, 763–765. One citation on page 6.

Kovacs, L., Zimmermann, A., Brockmann, G., Baurecht, H., Schwenzer-Zimmerer, K., Papadopulos, N.A., Papadopoulos, M.A., Sader, R., Biemer, E. & Zeilhofer, H.F. (2006). Accuracy and precision of the three-dimensional assessment of the facial surface using a 3-d laser scanner. *IEEE Trans Med Imaging*, **25**, 742–54. 11 citations on pages 17, 37, 40, 41, 44, 57, 59, 60, 61, 64, and 101.

Luhmann, T., Robson, S., Kyle, S. & Hartley, I. (2006). *Close Range Photogrammetry, Principles, Methods and Applications*. Whittles Publishing. 7 citations on pages 22, 23, 24, 25, 26, 27, and 34.

Majid, Z., Chong, A.K., Ahmad, A., Setan, H. & Samsudin, A.R. (2005). Photogrammetry and 3d laser scanning as spatial data capture techniques for a national craniofacial database. *The Photogrammetric Record*, **20**, 48–68. 8 citations on pages 17, 18, 36, 38, 58, 59, 64, and 100.

Manning, M.A. & Hoyme, H.E. (2007). Fetal alcohol spectrum disorders: A practical clinical approach to diagnosis. *Neuroscience & Biobehavioral Reviews*, **31**, 230 – 238. 4 citations on pages 5, 6, 7, and 75.

May, P.A., Brooke, L., Gossage, J.P., Croxford, J., Adnams, C., Jones, K.L., Robinson, L. & Viljoen, D. (2000). Epidemiology of fetal alcohol syndrome in a south african community in the western cape province. *Am J Public Health*, **90**, 1905–12. 2 citations on pages 5 and 7.

May, P.A., Gossage, J.P., White-Country, M., Goodhart, K., Decoteau, S., Trujillo, P.M., Kalberg, W.O., Viljoen, D.L. & Hoyme, H.E. (2004). Alcohol consumption and other maternal risk factors for fetal alcohol syndrome among three distinct samples of women before, during, and after pregnancy: the risk is relative. *Am J Med Genet*, **127C**, 10–20. One citation on page 7.

May, P.A., Gossage, J.P., Brooke, L.E., Snell, C.L., Marais, A.S., Hendricks, L.S., Croxford, J.A. & Viljoen, D.L. (2005). Maternal risk factors for fetal alcohol syndrome in the western cape province of south africa: a population-based study. *Am J Public Health*, **95**, 1190–9. One citation on page 7.

May, P.A., Gossage, J.P., Marais, A.S., Adnams, C.M., Hoyme, H.E., Jones, K.L., Robinson, L.K., Khaole, N.C.O., Snell, C., Kalberg, W.O., Hendricks, L., Brooke, L., Stellavato, C. & Viljoen, D.L. (2007). The epidemiology of fetal alcohol syndrome and partial fas in a south african community. *Drug and Alcohol Dependence*, **88**, 259–271. 5 citations on pages 1, 5, 7, 73, and 74.

May, P.A., Gossage, J.P., Marais, A.S., Hendricks, L.S., Snell, C.L., Tabachnick, B.G., Stellavato, C., Buckley, D.G., Brooke, L.E. & Viljoen, D.L. (2008). Maternal risk factors for fetal alcohol syndrome and partial fetal alcohol syndrome in south africa: A third study. *Alcoholism: Clinical and Experimental Research*, **32**, 738–753. 2 citations on pages 7 and 18.

McIntyre, G.T. & Mossey, P.A. (2003). Size and shape measurement in contemporary cephalometrics. *Eur J Orthod*, **25**, 231–42. One citation on page 69.

MedEIM (2004). Fcs2: Portable 3d colour facial imaging system. MedEIM Limited, Devizes, Wiltshire, UK. One citation on page 54.

Meintjes, E.M., Douglas, T.S., Martinez, F., Vaughan, C.L., Adams, L.P., Stekhoven, A. & Viljoen, D. (2002). A stereo-photogrammetric method to measure the facial dysmorphology of children in the diagnosis of fetal alcohol syndrome. *Med Eng Phys*, **24**, 683–9. 7 citations on pages 8, 13, 36, 38, 61, 62, and 66.

Mielke, P.W. & Berry, K.J. (2007). *Permutation Methods: A Distance Function Approach*. Springer, 2nd edn. One citation on page 86.

Mikhail, E.M., Bethel, J.S. & McGlone, J.C. (2001). *Introduction to modern photogrammetry*. Wiley, New York ; Chichester. 2 citations on pages 23 and 26.

Moore, E.S., Ward, R.E., Jamison, P.L., Morris, C.A., Bader, P.I. & Hall, B.D. (2001). The subtle facial signs of prenatal exposure to alcohol: an anthropometric approach. *J Pediatr*, **139**, 215–9. 7 citations on pages 2, 8, 9, 15, 74, 75, and 94.

Moore, E.S., Ward, R.E., Jamison, P.L., Morris, C.A., Bader, P.I. & Hall, B.D. (2002). New perspectives on the face in fetal alcohol syndrome: what anthropometry tells us. *Am J Med Genet*, **109**, 249–60. 7 citations on pages 2, 8, 10, 15, 74, 75, and 94.

Moore, E.S., Ward, R.E., Wetherill, L.F., Rogers, J.L., Autti-R?m?, I., Fagerlund, s., Jacobson, S.W., Robinson, L.K., Hoyme, H.E., Mattson, S.N. & Foroud, T. (2007). Unique facial features distinguish fetal alcohol syndrome patients and controls in diverse ethnic populations. *Alcoholism: Clinical and Experimental Research*, **31**, 1707–1713. 12 citations on pages 7, 8, 15, 17, 18, 20, 74, 75, 94, 95, 96, and 98.

Mutsvangwa, T. & Douglas, T.S. (2007). Morphometric analysis of facial landmark data to characterize the facial phenotype associated with fetal alcohol syndrome. *Journal of anatomy*, **210**, 209–220. 13 citations on pages 2, 8, 13, 15, 16, 38, 66, 73, 74, 75, 95, 96, and 98.

Naidoo, S., Harris, A., Swanevelder, S. & Lombard, C. (2006). Foetal alcohol syndrome: a cephalometric analysis of patients and controls. *Eur J Orthod*, **28**, 254–261. 2 citations on pages 15 and 94.

Photometrix (2004). *Australis users manual*. Photometrix Pty Ltd, PO Box 3023, Kew, Vic 3101, Australia. One citation on page 31.

Ras, F., Habets, L.L., van Ginkel, F.C. & Prahl-Andersen, B. (1996a). Quantification of facial morphology using stereophotogrammetry–demonstration of a new concept. *J Dent*, **24**, 369–74. One citation on page 61.

Ras, F., Habets, L.L.M.H., vanGinkel, F.C. & Prahl-Andersen, B. (1996b). Quantification of facial morphology using stereophotogrammetry – demonstration of a new concept. *Journal of Dentistry*, **24**, 369–374. One citation on page 36.

Remondino, F. & Fraser, C. (2006). Digital camera calibration methods: considerations and comparisons. In Isprs, ed., *International Archives of Photogrammetry, Remote*

Sensing and Spatial Information Sciences,, vol. XXXVI, Dresden, Germany. 2 citations on pages 26 and 27.

Robinson, D.L., Blackwell, P.G., Stillman, E.C. & Brook, A.H. (2001). Planar procrustes analysis of tooth shape. *Arch Oral Biol*, **46**, 191–9. 2 citations on pages 69 and 70.

Sampson, P.D., Streissguth, A.P., Bookstein, F.L., Little, R.E., Clarren, S.K., Dehaene, P., Hanson, J.W. & Jr, G.J.M. (1997). Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology*, **56**, 317–26. 3 citations on pages 5, 6, and 8.

Slice, D. (2005). *Modern Morphometrics in Physical Anthropology*. Kluwer Academic/Plenum Publishers, New York. 5 citations on pages 15, 65, 67, 68, and 70.

Slice, D.E. (2007). Geometric morphometrics. *Annu. Rev. Anthropol.*, **36**, 261–281. 2 citations on pages 65 and 67.

Somia, N., Rash, G., Epstein, E., Wachowiak, M., Sundine, M., Stremel, R., Barker, J. & Gossman, D. (2000). A computer analysis of reflex eyelid motion in normal subjects and in facial neuropathy. *Clinical Biomechanics*, **15**, 766–771. One citation on page 38.

Stallings, W.M. & Gilmore, G.M. (1971). A note on "accuracy" and "precision". *Journal of Educational Measurement*, **8**, 127–128,129. One citation on page 59.

Stegmann, M.B. & Gomez, D.D. (2002). A brief introduction to statistical shape analysis. Technical University of Denmark, Lyngby. One citation on page 68.

Streissguth, A.P., Aase, J.M., Clarren, S.K., Randels, S.P., LaDue, R.A. & Smith, D.F. (1991). Fetal alcohol syndrome in adolescents and adults. *JAMA*, **265**, 1961–7. One citation on page 66.

Swiderski, D.L. (2003). Separating size from allometry: Analysis of lower jaw morphology in the fox squirrel, sciurus niger. *Journal of Mammalogy*, **84**, 861–876. One citation on page 81.

Tobin, J.L., Di Franco, M., Eichers, E., May-Simera, H., Garcia, M., Yan, J., Quinlan, R., Justice, M.J., Hennekam, R.C., Briscoe, J., Tada, M., Mayor, R., Burns, A.J., Lupski, J.R., Hammond, P. & Beales, P.L. (2008). Inhibition of neural crest migration underlies craniofacial dysmorphology and hirschsprung's disease in bardetbiedl syndrome. *Proceedings of the National Academy of Sciences*, **105**, 6714–6719. One citation on page 54.

Viljoen, D., Croxford, J., Gossage, J.P., Kodituwakku, P.W. & May, P.A. (2002). Characteristics of mothers of children with fetal alcohol syndrome in the western cape province of south africa: a case control study. *J Stud Alcohol*, **63**, 6–17. One citation on page 7.

Viljoen, D., Craig, P., Hymbaugh, K.., Boyle, C. & Blount, S. (2003). Fetal alcohol syndromesouth africa, 2001. *Morbidity and Mortality Weekly Report*, **52(28)**, 660–662. One citation on page 103.

Viljoen, D.L., Gossage, J.P., Brooke, L., Adnams, C.M., Jones, K.L., Robinson, L.K., Hoyme, H.E., Snell, C., Khaole, N.C., Kodituwakku, P., Asante, K.O., Findlay, R., Quinton, B., Marais, A.S., Kalberg, W.O. & May, P.A. (2005). Fetal alcohol syndrome epidemiology in a south african community: a second study of a very high prevalence area. *J Stud Alcohol*, **66**, 593–604. One citation on page 7.

Webb, A.R. (2002). *Statistical Pattern Recognition*. John Wiley and Sons Ltd. One citation on page 78.

Weinberg, S.M. & Kolar, J.C. (2005). Three-dimensional surface imaging: limitations and considerations from the anthropometric perspective. *The Journal of craniofacial surgery*, **16**, 847–851. One citation on page 18.

Weinberg, S.M., Scott, N.M., Neiswanger, K., Brandon, C.A. & Marazita, M.L. (2004). Digital three-dimensional photogrammetry: evaluation of anthropometric precision and accuracy using a genex 3d camera system. *The Cleft palate-craniofacial journal*, **41**, 507–518. 7 citations on pages 17, 40, 41, 44, 64, 100, and 101.

Weinberg, S.M., Naidoo, S., Govier, D.P., Martin, R.A., Kane, A.A. & Marazita, M.L. (2006). Anthropometric precision and accuracy of digital three-dimensional photogrammetry: comparing the genex and 3dmd imaging systems with

one another and with direct anthropometry. *J Craniofac Surg*, **17**, 477–83. 8 citations on pages 17, 40, 42, 59, 60, 61, 62, and 64.

Winder, R., Darvann, T., McKnight, W., Magee, J. & Ramsay-Baggs, P. (2008). Technical validation of the di3d stereophotogrammetry surface imaging system. *British Journal of Oral and Maxillofacial Surgery*, **46**, 33–37. 3 citations on pages 40, 59, and 60.